

基于规则和案例推理集成的刑事 案件量刑决策支持研究

余贵清 著

電子工業出版社

Publishing House of Electronics Industry

北京 · BEIJING

内 容 简 介

本书构建了基于本体的审判案例自动抽取标注模型；针对审判案例特点，提出了支持多属性、多案例、多要素的 GA-KNN 案例检索方法；利用决策树、神经网络等数据挖掘算法构建了规则推理和案例推理集成的决策支持机制。对以半结构化或非结构化数据作为决策基础数据，以规则推理和案例推理同时支持的决策工作具有重要借鉴意义。

未经许可，不得以任何方式复制或抄袭本书之部分或全部内容。
版权所有，侵权必究。

图书在版编目（CIP）数据

基于规则和案例推理集成的刑事案件量刑决策支持研究 / 余贵清著. —北京：电子工业出版社，2016.3

ISBN 978-7-121-28142-6

I. ①基… II. ①余… III. ①刑事犯罪—量刑—研究—中国 IV. ①D924.134

中国版本图书馆 CIP 数据核字（2016）第 026454 号

责任编辑：秦绪军

特约编辑：田学清 赵海红

印 刷：

装 订：

出版发行：电子工业出版社

北京市海淀区万寿路 173 信箱

邮编：100036

开 本：720×1000 1/16 印张：16 字数：258 千字 彩插：1

版 次：2016 年 3 月第 1 版

印 次：2016 年 3 月第 1 次印刷

定 价：45.00 元

凡所购买电子工业出版社图书有缺损问题，请向购买书店调换。若书店售缺，请与本社发行部联系，联系及邮购电话：（010）88254888。

质量投诉请发邮件至 zltz@phei.com.cn，盗版侵权举报请发邮件到 dbqq@phei.com.cn。

服务热线：（010）88258888。

前言

随着国内经济发展、社会进步，国内司法领域审判量刑决策的科学性、规范性成为社会关注的焦点，法规与案例结合支持法官量刑决策成为司法审判领域和管理理论界的热点问题，迫切需要进行针对性、前瞻性、应用性强的研究。为了给法官决策提供有效的支持，本书创建了基于规则和案例推理集成的辅助量刑决策机制，并运用本体、遗传算法等理论与方法，结合案例进行了严谨的论证。本书研究主要构建了基于本体的刑事案例表示方法，提出和论证了对于多属性、多案例、多要素背景下，基于 GA-KNN 的案例属性优化与检索针对性方案，揭示了基于规则与案例推理集成的量刑决策机制。为了系统论证集成案例推理的可靠性，以 11 000 个故意伤害罪的半结构化案例文档集为初始样本，通过本体技术进行决策信息规则化、结构化，采用 GA-KNN 算法进行属性优化与案例检索，采用决策树和神经网络模型构建推理集成决策模型，通过与实际量刑决策结果比较论证了该方法对支持量刑决策有重大作用。

本书主要研究内容和结论归纳如下：

本书针对目前法官量刑决策判决书文档集的非结构化、半结构化特征，首先，通过分析刑事案件判决书文档集结构特征，从本体理论规范知识表达出发，建立了刑事案件本体，并以罪名为单位，对刑事案件相关的概念、类和属性进行了规范化表达；其次，以案例文档集和刑事案件本体为基础，通过信息抽取

和标注技术,建立了半结构化文本决策数据的自动抽取模板,实现了刑事案件实体信息的自动化描述性抽取;再次,通过数据仓库技术,对刑事案件决策信息进行逻辑设计,将描述性实体信息规则化,并形成规则化案例库;最后,对刑事案件本体设计和信息抽取方法进行了验证。研究发现,以本体为核心,从模型、实验和评价三个角度阐述的从半结构化数据向结构化案例库转化的机制,是利用信息技术手段实现更多决策信息可利用的重要途径,是科学决策的重要基础。该研究结论确立了蕴含在半结构化文档中决策信息支撑决策的途径,突破了从半结构化数据中获取决策信息的瓶颈,实现了本体理论在领域决策中的应用。

本书针对司法领域的多案例、多属性、多要素特征及推理机制依赖的案例库必须满足快速检索、及时支持法官决策的要求,确立了刑事案例属性优化与敏捷检索这一关键技术问题。首先,通过刑事案例的特点分析确立了刑事案例推理精确性目标与案例检索环节方法选择与路径;其次,通过遗传算法、改进的 AHP 方法对案例属性进行优化效果比较,形成适合刑事案件量刑决策的属性优化方法,分析比较了不同方法检索结果异同;最后,论证了多案例、多属性、多要素司法类案例应用 GA-KNN 方法组合完整检索方法的可行路径与检索精度。研究发现,基于权重优化的 GA-KNN 检索方法可以提高多案例、多属性、多要素司法类案例的检索精确度及检索效率。该研究拓宽了管理科学与信息技术交叉研究的范围,实现了决策过程依赖特定决策基础信息的最优集合,使决策推理可以在有限的时间内完成。

本书针对我国成文法依法规量刑和结合案例辅助约束法官自由裁量权的需要,重点研究刑事案件量刑决策中规则推理和案例推理集成决策机制问题。首先,分析了以案例库为基础的数据挖掘技术,根据刑事案件信息离散性和决策过程的非线性特点,提出并论证了决策树模型和神经网络模型的综合应用模式,

使法规与案例在决策推理中进行结合；其次，利用实验方法，对两种推理集成的模型进行训练，直至满足决策要求，将推理模型按案件类别（罪名）进行确定，形成各类型案件对应的决策模型；最后，通过具体使用该推理集成决策模型，对其应用效果进行了验证。研究发现，利用数据挖掘方法对领域知识进行深度挖掘，可以找到决策中潜在的关键信息，如量刑决策特征中的暴力方式细节与决策结果的关系、量刑结果类型与各犯罪特征集的对应等。该研究结论中两种推理的集成充实了管理决策方法，同时，拓宽了领域知识的挖掘和利用机制。

本书针对通过信息决策系统自动支持法官量刑的需要，确定了如何将决策模型在实际量刑决策中得以应用的前瞻性研究问题。首先，从上述研究成果出发，从系统需求与功能模块划分的角度对系统进行了分析与阐述，提出了量刑决策辅助系统原型；其次，基于角色和目标案例，进行了系统架构设计，研究了三个关键模块的设计与实现细节；最后，对该系统原型进行了初步实现，并展示了相关界面，验证了该原型的正确性。研究发现，建立量刑决策辅助系统需要有强大的数据库、知识库做支撑，将案件的案情结构化、法官化是其中的重要环节。该研究结论是决策理论、信息技术与领域知识等多学科研究成果综合应用的结果，拓宽了多学科交叉应用研究的领域。

关键词：量刑决策；刑事案件本体；案例属性；案例推理；推理集成

目 录

第 1 章 绪 论	1
1.1 研究背景及问题提出	1
1.2 研究意义	3
1.2.1 理论意义	4
1.2.2 实践意义	6
1.3 本书中相关概念界定	7
1.3.1 半结构化文本	7
1.3.2 领域本体	7
1.3.3 信息抽取	8
1.3.4 语义标注	9
1.3.5 刑罚裁量	10
1.3.6 案例推理	10
1.3.7 决策集成	11
1.4 研究对象与研究内容	11
1.4.1 研究对象	11
1.4.2 研究内容	11
1.5 研究方法与技术路线	13
1.5.1 研究方法	13
1.5.2 技术路线	14

1.6 章节安排	16
1.7 研究的创新之处	18
第 2 章 国内外研究综述	20
2.1 法律知识的表示与抽取	20
2.1.1 法律知识的表示	21
2.1.2 法律知识的抽取	28
2.2 法律案例的属性优化与案例检索方法	38
2.2.1 案例属性优化	38
2.2.2 案例检索方法	41
2.3 法律知识的推理	42
2.3.1 法律知识的推理逻辑	42
2.3.2 法律知识的推理方法	44
2.4 文献述评	50
2.4.1 研究问题的价值	50
2.4.2 研究面临的问题	52
第 3 章 基于本体的刑事案例库构建研究	56
3.1 构建刑事案例库的条件分析	57
3.1.1 刑事案件文书特点	57
3.1.2 刑事案例现状分析	59
3.1.3 刑事案件本体构建难度	60
3.2 刑事案件本体化框架设计	62
3.2.1 本体建模基本模型	62
3.2.2 刑事案件本体构建	64
3.2.3 刑事案件信息抽取	74
3.2.4 刑事案例库设计	81
3.3 刑事案件本体构建实现与信息抽取评价	85
3.3.1 刑事案件本体构建实现	85

3.3.2 刑事案件信息抽取评价	88
3.4 本章小结	91
第 4 章 基于遗传算法的刑事案例属性优化与检索策略研究	92
4.1 基于遗传算法对刑事案例属性优化的研究.....	93
4.1.1 刑事案例属性的特征及属性优化的意义	93
4.1.2 两种优化方法对刑事案例属性优化的针对性分析	94
4.1.3 针对刑事案例属性优化的两种优化方法对比分析	97
4.2 基于 GA-KNN 算法对刑事案例检索策略研究	108
4.2.1 案例相似性的 KNN 算法	108
4.2.2 案例相似性的 GA-KNN 算法	112
4.2.3 案例相似性的 KNN 方法与 GA-KNN 算法的比较	113
4.3 本章小结	116
第 5 章 刑事案件推理集成决策模型研究	117
5.1 基于本体的法律推理体系	117
5.2 刑事案件推理决策模型中的相关算法	119
5.2.1 特征选择	119
5.2.2 决策树	123
5.2.3 神经网络	128
5.3 刑事案件推理集成决策模型研究	130
5.3.1 推理集成决策模型建立意义	130
5.3.2 推理集成决策模型设计	132
5.3.3 案例特征提取策略	137
5.3.4 模型训练阶段的方案设计	139
5.3.5 模型使用阶段的方案设计	140
5.4 刑事案件推理集成模型实现	141
5.4.1 实验平台	141
5.4.2 实验数据及预处理	141

5.4.3 判决类型推理模型实验	146
5.4.4 判决刑期推理模型实验	150
5.5 刑事案件推理集成模型评价	151
5.6 本章小结	152
第 6 章 刑事案件量刑决策辅助系统原型设计	154
6.1 需求分析	155
6.1.1 系统需求概述	155
6.1.2 系统功能需求	155
6.2 系统总体设计	156
6.2.1 系统工作流程	157
6.2.2 系统架构设计	159
6.3 功能模块实现	160
6.3.1 案例库模块实现	160
6.3.2 决策推理模块实现	161
6.3.3 案例推荐模块实现	164
6.4 系统原型界面设计	165
6.5 本章小结	168
结论与展望	170
附录 A 刑事案件本体类与属性的详细设计	173
附录 B 刑事案件本体 OWL 文件	180
附录 C 故意伤害罪类刑事案件本体	215
附录 D 刑事案件本体 XML 抽取模板	219
附录 E 实验数据	230
后 记	233
参考文献	235

绪 论

1.1 研究背景及问题提出

伴随着我国经济飞速发展和人民生活水平不断提高，人们对社会公平正义的呼声越来越高，依法治国成为我国政治发展的必然趋势。2010年，我国的社会主义法律体系初步建成，明确要求法官遵循法律原则和法律规定的情况下，准确掌握案件事实，慎重把握裁判尺度，妥善处理矛盾纠纷，实现社会公平正义。

量刑规范化和案例指导制度等成为法律理论界与实务界关注的重点，迫切需要通过信息技术手段解决量刑不统一的问题，为法院科学决策、科学发展进一步提供理论技术支持。因为我国法律体系建立时间不长，存在着两个天生的不足之处：一是我国唯一的量刑主体——法官队伍中，存在着大量理论功底缺乏和审判经验不足等问题，表现为从事刑事审判的法官不能很好地驾驭审判，

对案情不能形成较全面的自由心证，对如何量刑不能做出正确判断。二是《刑法》规定宽泛不严谨，使法官自由裁量权过大，成为司法不统一的主要因素。世界量刑体制大体有立法模式、司法模式、行政模式等。我国目前采用的是司法模式，也就是法律上对犯罪行为规定了相对确定的幅度刑，法官对具体案件宣判的是确定刑（定期刑）。我国的刑事立法规定的每个罪，应受的处罚要么过于宽泛，要么过于笼统。例如，《刑法》第 234 条规定，犯故意伤害罪“致人死亡或者以特别残忍手段致人重伤造成严重残疾的，处十年以上有期徒刑、无期徒刑或者死刑”，到底什么情况判多少年、什么情况判死刑由法官作判。为解决上述司法实践中的缺陷，近几年，最高法院连续发布量刑规范化办法，并着手建立案例指导制度，两个重要课题都离不开信息技术的支持，需要将管理决策理论、司法审判领域知识和信息技术相结合做前瞻性研究。

人民法院信息化的不断发展，为建立基于案例指导的司法审判制度提供了海量的决策数据，为科学决策提供了数据基础。我国法院信息化发展大都经历了十几年的快速发展，不仅有支持法官办案的审判业务系统，同时，也初步探索建立了支持决策的系统，并积累了数以千万计的电子化判决书。这些判决书蕴含着案件的案情、判决依据、判决结果和审理理念，它具有数据量大和半结构化两个特点。由于判决书量大，具有大数据体量巨大、数据类型繁多、价值密度低和处理速度快的特征，不可能采用人工方法，将其转成支持量刑决策的信息；由于其半结构化的特点，经过恰当的方式进行结构分析和表达，可以将其转化为结构化的决策信息。

现代决策理论的发展，使决策理论吸收了行为科学、系统理论、信息技术及运筹学等新兴学科的理论和方法，产生了决策技术论分支。决策技术论认为在决策过程中要对各种各样的决策进行深入细致的分析，借助一定的科学方法和手段，发展相应的决策技术，可以减少决策主体有限理性对决策结果的影响。

支持大数据的数据库、云计算、可扩充存储系统和挖掘技术的不断发展,使得将有价值的决策信息从半结构化的案例文本中进行挖掘,供决策中使用成为可能。数据库和数据挖掘的产生和发展,使得当今的计算机网络应用体系从业务管理逐步跃升到决策支持。同时,在“大数据”时代,要从海量、纷繁复杂的数据中快速找到决策者关注的信息,必须借助功能强大、智能化的信息检索工具。将系统蕴含的基础数据转化为决策知识,也为大规模数据统计分析处理提供了一套完整的解决方案。

面对社会对法律需求的机遇和法官努力实现司法统一的内在动力,从最高法院到地方法院都着手研究,借助网络与信息技术,就基于规则和案例支持的新的量刑决策模式进行前瞻性研究,创新提出量刑决策新模式。从而,提高司法审判的科学性和量刑的均衡统一,使审判决策结果朝着“同案同判”的方向发展。为此,从决策科学的角度入手,探索信息条件下的决策方法论,支持在审判领域创建规则和案例推理集成的决策模式,以建立一套相对科学的、处理复杂决策问题行之有效的支持动态决策的方法,这是本书研究的出发点和落脚点。

1.2 研究意义

在司法审判领域,我国采用的是成文法,坚持“有法可依”的原则,刑事案件审判的基本原则是以事实为根据,以法律为准绳,进行量刑决策。根据案件性质的不同,我国《刑法》将刑事案件分为400多个罪名,每个罪名又根据案件事实的区别,给出一定的量刑幅度,具体刑期就由法官进行自由裁量。

法官自由裁量的依据有三个，一是依据“重现的”案件事实。刑事案件审理都是事后的，依据的案件事实是通过各种审理的必要程序进行“重现的”，其带有明显的与真实事实的差异。同时，实践表明，即使相同罪名的案件，其情节也千差万别，各不相同，如何找出各类案件的共同特征，理清案情的关键要素，用以表示相同或相似案件的特征是从整体研究审判决策方法的前提。二是依据相关法律。我国法律体系初步建立，加上社会环境日新月异，法律条文的宽泛难以适应新的社会矛盾和审判工作的需求。以故意伤害罪在《刑法》第 234 条中的表述为例，“故意伤害他人身体的处三年以下有期徒刑、拘役或者管制。犯前款罪，致人重伤的，处三年以上十年以下有期徒刑；致人死亡或者以特别残忍手段致人重伤造成严重残疾的，处十年以上有期徒刑、无期徒刑或者死刑。”从中可见，无论从法的规定，还是量刑区间，都存在很大的弹性，需要裁判主体通过智慧去合理裁判。三是依据法官的智慧。法官的智慧和经验是隐性知识，来源于对大量相同或相似案件的审理，以及对法律法规的反复解释，是一种实践和参悟的结果。不能否认有许多优秀和成熟的法官具有这样的能力，但绝大多数法官并不具备这样的能力和素质，需要借助成熟的决策方法的指引。

以大量的案件数据为依据，通过科学的数据抽取方法，挖掘案件量刑的规则，建立案例库，形成辅助审判决策的信息平台，无论是在决策理论研究，还是决策方法实践研究，都具有重要的意义。

1.2.1 理论意义

决策技术论，即对各种各样的决策进行深入细致的分析，提出有针对性的方法，发展相应的决策技术。决策理论学派所主张的决策，是在解决复杂决策问题时，要求尽可能量化，建立数学模式，从而求得满意结果。即使结果满意，

计算工作量也是很大的,离开了电子计算机有时是很难办到的。计算机科学的兴起是决策理论学派得以存在和发展的基础。本书正是从帮助决策者处理信息、解决关键技术、建立决策模型并提供可计算机系统化的决策方法的研究,属于决策理论学派范畴。

目前,不少学者在不同领域研究规则推理,也有不少应用成果。案例推理由于其案例表达的困难和长期以来信息挖掘的瓶颈,使其处在研究较多,但专业领域应用却不是很深入的境地。将规则和案件推理相结合,应用到司法审判领域就更少见。这就需要管理科学与工程领域的决策管理研究者,对决策活动中的这些信息从挖掘技术、智能集成和辅助决策等方面进行研究,为决策者提供一套处理复杂决策问题的理论方法和智能决策途径与手段,形成信息条件下的决策模式,用以指导决策活动^[1]。

(1) 决策信息的表达方式。以本体理论为基础,依据刑事案件自身特征,探索领域知识本体化表示方法,为使半结构化或非结构化的决策信息规则化和利用化成为可能,拓宽了决策信息的规则化表达方式。

(2) 信息技术为决策服务。以数据挖掘技术为基础,对刑事法律法规和刑事案例进行本体化标注和实体信息的抽取研究,并以此形成支持决策的案例库,使决策活动直接依赖大数据提高决策水平,给管理科学的发展增加了新的内容。

(3) 推理集成促进决策质量。探索在专业领域,进行规则推理和案例推理的集成,形成从不同角度保证科学决策,提出新的决策路径。信息理论与决策理论的结合,会在不同领域的决策中起到示范效应,为其他相关决策研究提供新视角和方法。

(4) 管理科学与相关领域的交叉研究。将管理科学中知识表示、数据挖掘和决策方法与刑事案件量刑中的案件要素、法律法规推理和参阅案例研究进行了很好的结合,对服务专业领域决策具有一定的借鉴意义。

1.2.2 实践意义

法官作为量刑决策的主体,面临着来自知识、能力、领导和社会的各种压力,法官群体迫切需要用技术方法帮助其解决量刑决策问题,而且希望这种决策支持是自动的。通过数据挖掘和推理技术,在刑事案件量刑过程中具体解决拥有大量数据而有用信息却很贫乏的这种不正常的现象,无疑具有很强的现实意义。

(1) 确定描述刑事案件知识的方法。分析大量刑事案件,从刑事案件的组成细化和归纳出其构成,为法官提供一个分析案情的路线图,是绝大多数法官,特别是年轻法官迫切需要的。结合案情与裁判结果,努力借助大数据分析的思想,将法官从深陷于因果分析的劳累中解放出来,利用相关关系进行逻辑推理与判断,可以提高审判效率和审判质量,减轻审判的压力。

(2) 支持量刑决策的推理集成决策方法。法规与案例的数字化,使法官更容易得到法规的支持和案例的帮助。根据目标案件的案情,自动确定案件的刑罚类型,通过刑罚类型选择案例支持,进而得到目标案件的量刑结果和支持量刑的参考案例,使判决结果趋于均衡化。

(3) 提高量刑决策的能力。基于文本挖掘技术的发展,使得对法律法规和蕴含法官量刑智慧的判决书得以有效的信息挖掘,有效地提高了法官智慧的积累速度,扩大了法官经验传承的范围。从知识管理角度讲,提高了决策者内在的决策水平,突破了支持决策的知识瓶颈。

总之，研究案件规则化和推理集成，从而支持量刑决策，是对决策理论的丰富，也是决策理论在司法审判领域的创新应用，具有重要的理论价值与实践意义。

1.3 本书中相关概念界定

1.3.1 半结构化文本

半结构化文本是指满足以下条件的文本：文本由具有独立语义的内容模块组成，各模块可以由若干名词组成；每个独立模块间由非标点分隔，如回车；文本各模块内容可以归纳为一个名词集合，所有同类文本对应的集合是一个确定的有限集；每个模块内部语义的描述不一样，如简历和公文。法院的判决书是典型的半结构化文本。

1.3.2 领域本体

本体是用来描述某一领域甚至更广范围内的概念定义及概念之间的联系，使得这些概念和联系在共享的范围内有明确唯一的定义，达成一种共识，方便人机交流。领域本体是指在通用本体的基础上，以一个特定领域为描述对象的本体，提供该特定领域的概念定义和概念之间的关系、主要理论和基本原理，以及领域中发生的活动等。法律领域的本体可以分为法律条例本体和法律案例本体，刑事案件本体包括刑事法规本体和刑事案例本体。本书着重研究在对新案件进行量刑时，可以根据法规本体检索出相似案例进行刑罚类型的判定，通过案例本体进行刑期的判定。刑事案件本体属于法律领域本体的概念范畴。

1.3.3 信息抽取

信息抽取分为广义上的信息抽取和狭义上的信息抽取。狭义上的信息抽取是指从文本中抽取指定的事件、事实等信息，形成结构化数据，存入数据库或语义库中，供法官进行更人性化的查询和使用的过程。信息抽取的实质是从文本中抽取出法官感兴趣的事件、实体和关系，并将其进行结构化描述、结构化存储，为信息检索、文本分类等其他应用提供服务。广义上的信息抽取技术主要是对抽取对象的扩展，即将其他形式的信息作为抽取对象，如图像、音视频等多媒体信息，抽取结果为相应的结构化数据。除此之外，还有伴随 Internet 中的海量信息出现的 Web 信息抽取，即从基于网页的半结构化的 Web 文档中提取信息点^[2]。本书信息抽取为狭义上的信息抽取。

信息抽取的过程可以理解为从待处理文本中抽取信息，然后依次填入输出模板相应的槽中。输出模板由多个槽组成，它是信息抽取系统结构化的输出结果。Cardie C^[3]曾给出信息抽取的通过程，如图 1-1 所示。

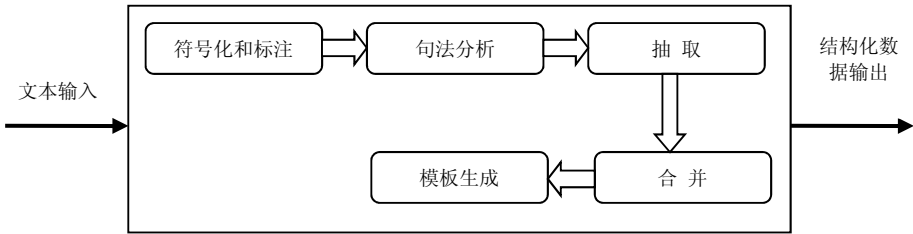


图 1-1 信息抽取过程

(1)符号化和标注主要用于对输入数据分段、分句或分词后进行词性标注，有些系统还会加入语义标注。

(2)句法分析主要用来区分文本中的名词短语、动词短语等各种语法结构，从中识别出与抽取任务相关的各类命名实体。

(3)抽取利用与领域相关的抽取模式来识别文本中各个命名实体间的关系,然后根据具体抽取任务将需要的信息抽取出来,填入到输出模板的槽中。

(4)合并用于解决文本中命名实体的共指消解问题,如果发现两个指代指向同一个命名实体,则将两个指代合并。

(5)模板生成主要完成推理和新模板生成的工作,推理是根据具体抽取任务并结合领域知识来对文本进行推断,以得出目标抽取信息。

1.3.4 语义标注

语义标注^[4]是以特定领域本体为指导,为领域文档集添加规范化知识表示的过程,即在定义好一个由 OWL (Ontology Web Language, OWL) 语言书写的本体后,将文档中文本知识用 RDF (Resource Description Framework, RDF) 语言描述出来。这个过程通常可以分为两个步骤:

(1) 类型标注。将文档集中与本体中概念相对应的词语标记出来,作为该概念所对应的实例,通常用 RDF 资源的形式来表示。

(2) 关系抽取。找出实例之间存在的与本体中属性相对应的关系,一般将相互关联的两个实例及实例之间的关系表示为 $(R1, P, R2)$, 即一个 RDF 陈述。 P 表示两个实例 $R1$ 和 $R2$ 间的关系,对应本体中的一个属性。

目前已有的关于语义标注的工作主要包括三种方法:一种是基于机器学习和关系元数据的传统信息抽取与标注技术,主要是在标注好的训练集上进行训练。这种方法不能有效地利用本体中已有的信息,在实际工作中,需要对匹配规则进行自行定义或使用其他方法来实现抽取结果向本体内知识的转换。另一

种是基于本体的信息抽取与标注方法,利用本体中已有的实例来简化抽取过程中对概念实例的识别。如 Dill S 等^[5]在 TAP 本体实例集合中查找所有与待标注词相匹配的实例集合,然后通过比对待标注词语的上下文与实体中实例的上下文,构造相应的文本向量,进行相似度计算,找到与待标注词最相匹配的实例。但在这种方法中,主要是对文本中实例的抽取和标注,对实体间关系的抽取力度较弱,这就需要第三种方法:基于自然语言处理的方法,借助对语法结构和句法结构的分析,进行实体关系抽取。在实现实体识别的工作后,通过谓语句词进行本体属性的映射,如 Alani H^[6]。Schutz A 等^[7]运用统计方法获取领域中表达属性关系的动词,并在句法分析过程中运用这些动词来寻找对应的本体。Tenier S 等^[8]提出最小子树的概念,利用 DOM(Document Object Model, DOM)树结构来辅助关系抽取。

1.3.5 刑罚裁量

刑罚裁量简称量刑,指人民法院根据行为人所犯罪行及刑事责任的轻重,在定罪并找准法定刑的基础上,依法决定犯罪分子是否判处处罚、判处何种刑罚的审判活动。量刑包括刑罚类型和具体刑期。本书中的刑罚类型是在法律规定基础上,根据案情将其进一步细化的结果。

1.3.6 案例推理

案例推理是指当遇到一个新的问题时,系统根据关键的特征在原始案例库中进行检索,找出一个与待求解问题最相近的候选案例或案例集,重用此候选案例或案例集的决策方法。

1.3.7 决策集成

决策集成指在同一决策过程中将多种决策推理方法或推理结果进行相结合,目的是使决策更加科学化。

1.4 研究对象与研究内容

1.4.1 研究对象

本书以《刑法》和刑事案件起诉书、判决书为研究对象,通过《刑法》对罪名类型的分类和特征描述,作为刑事案件的分类标准和法规本体库的索引,罪名及犯罪情节特征及对应的量刑幅度作为检索内容。以××法院近10年来,积累的140万件刑事案件的起诉书、判决书作为案件特征提炼、归纳的基础数据,按不同罪名及犯罪程度,进行影响量刑的指标因素重要性分析,建立集中统一的案例数据库,为决策研究提供数据基础。本书以故意伤害罪为例,选择了11 000个判决书作为研究对象。

1.4.2 研究内容

长期以来,我国刑罚裁量的模式是人民法院根据行为人所犯罪行情节及刑事责任轻重,在定罪并确定法定刑的基础上,依法决定对犯罪分子是否判处处罚、判处何种刑罚。由于其中多种因素的不确定性和决策者认知的不同,众多决策主体无法做到对相同或相似犯罪事实做出相同的裁量结果。本书力图从基于案例的角度,探索刑罚裁量新模式,用技术手段辅助规则裁量,规避现实遇到的问题。

本书主要包括以下研究内容：

（1）针对支持法官量刑决策的案例是以半结构化的判决书文档集存在的这一现实，确立了采取何种规范化描述方式才能更好地创建可推理的决策数据资源这一问题。通过文档结构分析和本体理论相结合，实现判决书文档集的规则化、结构化表示，支持关键决策基础信息的抽取和案例库的创建。

（2）针对推理机制依赖的案例库必须满足快速检索、及时支持法官决策的要求，确立了刑事案例属性优化与检索问题。案例属性的优化与检索，是为法律推理准备案例库的前提，只有用恰当的案例对推理模型进行训练，其在解决目标案例中的量刑问题时才有效。

（3）针对我国成文法依法规量刑和结合案例辅助约束法官自由裁量权的需要，确定了探析在刑事案件量刑决策中规则推理和案例推理集成决策机制的问题。将推理过程集成化，创建自动支持法官量刑决策的模式，是本书研究的目标。

（4）针对通过信息决策系统自动支持法官量刑的需要，确定了如何将决策模型在实际量刑决策中得以应用的前瞻性研究问题。在强大数据库的支持下，提出面向决策主体的支持决策辅助系统原型设计，使本书研究成果得以实现。

上述问题，由于决策途径的探索不够和决策信息的瓶颈，致使我国刑事案件的审判还停留在人对案件事实和量刑标准的自我感知上，还没有充分利用现存的大量半结构化的案例文本。这一决策模式的构成、决策知识的获取和决策方法的结合都处在尝试阶段，需要更多管理科学方面的专家的支持。本书将针对上述一系列问题设计创新点，展开研究。

1.5 研究方法与技术路线

1.5.1 研究方法

1. 案例研究方法

厘清刑事案件构成要素，并用规范的方法表达出来，是本书的主要研究内容之一。鉴于××法院信息化发展起步早、基础好，刑事案件也是社会反应比较强烈的案件，对研究司法审判领域决策具有典型性和代表性，所以以××法院刑事案件审理为案例，研究信息条件下的量刑决策模式及辅助决策模型是一个较好的选择。

2. 实证研究方法

针对刑事案件量刑决策中重要依赖的起诉书和历史案例，能否将影响决策的信息抽取出来，它是决定决策模型建立的基础信息问题，本书采用了实证研究方法。首先，在文献研究和理论分析的基础上，提出了基于本体的案例实体抽取和关系抽取模型，然后通过对“故意伤害罪”实际案例的抽取，获取决策数据，支持推理模型的建立。分析时所采用的软件工具主要包括：Protégé 4.1、Sybase IQ 和 SPSS Clementine。

3. 数据挖掘方法

针对半结化或非结构化的案件信息，通过数据挖掘方法进行实体信息的提取，从而形成可用于决策的数据库。本研究主要用了数据仓库技术、数据挖掘技术，以及推理决策中的神经网络和决策树技术，将其中的技术应用到领域决策中，使决策活动得到更大的信息支持。但本书只是拿来应用，而没有就同一应用，提供不同的技术手段并加以比较进行研究。

1.5.2 技术路线

本书借鉴国内外相关研究成果，以智能决策理论、信息抽取理论、大数据思想为依托，总结出大数据背景下的智能决策研究的总体思路和关键技术，建立基于规则和案例推理集成的刑事案件智能决策辅助系统的研究框架。依据研究框架，重点研究了刑事案件本体设计与建立、基于本体的刑事案例信息抽取、规则和案例推理集成决策模型和刑事案件量刑决策系统原型设计。

总体上分，研究方案分为两部分，一是案件本体及本体案例库的设计与建立，二是对目标案例的处理过程，即案例推理方法进行研究。另外，设计并实现刑事案件案例推理系统，面向审判人员提供相关服务与数据支持。

1. 刑事案件案例库的建立

如图 1-2 所示，刑事案件案例库的建立可分为四个过程：基于本体建立刑事案件本体模型、刑事案件本体建立、以刑事案件本体为模板进行刑事案件信息抽取、建立刑事案例库。

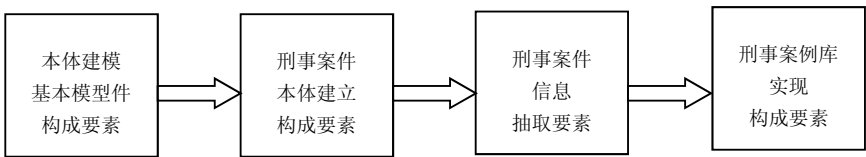


图 1-2 刑事案件案例库建立过程

2. 目标案例处理的过程

如图 1-3 所示，目标案例的处理可分为四个阶段，包括预处理、特征提取、特征匹配、结果反馈。本研究考虑设计并建立刑事案件案例推理系统原型对上述案例处理过程进行支持。

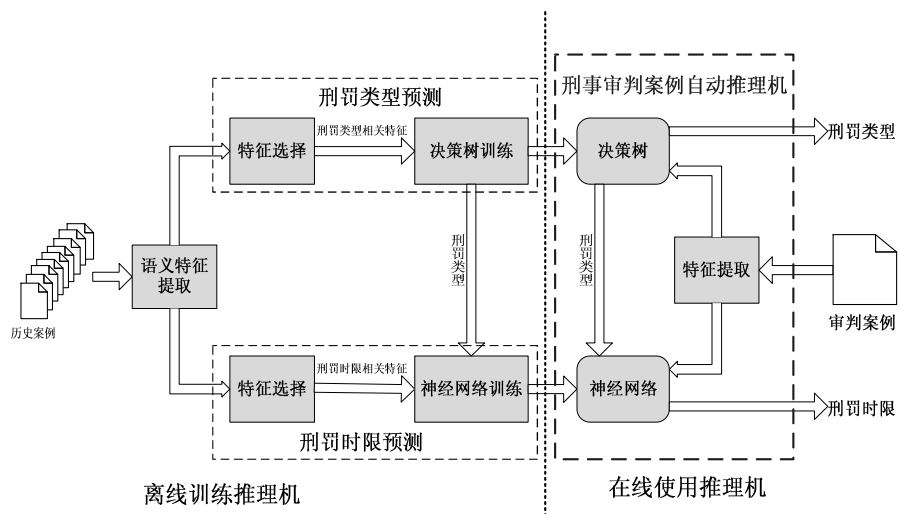


图 1-3 目标案例处理的过程

1) 待处理案例的预处理

这一阶段处理过程的作用主要是对案例文本进行预处理。首先，对于文本中较为明显且易被自动化识别的语法、拼写错误进行修正；其次，将文本的格式调整为统一的格式以方便后期的处理，若原始文本中存在多余的内容也需在这一步中予以删除，以增加后面其他处理过程的效率和准确程度；然后，对文本进行分词和词性标注，在分词和词性标注的过程中，需要基于目前较为成熟的中文分词和词性标注算法及领域词典等。这样，就得到了一个较为纯净且有助于后续处理的本书文档。

2) 待处理案例的案例特征提取

这一阶段的主要目的是找到待处理文档的特征，即待处理案例所涉及的场景与所提到事实、证据、事件细节等。基于案例文档半结构化的特点，结合领域词典的协助，从两个方面提取其特征：针对案例文档中的结构化部分，提取其概念框架；针对非结构化部分，基于案例文档的书写、表述特征提取场景信

息、属性和事件细节等。最后，将上述两方面的信息结合，整理为 XML 格式的案例特征文档。

3) 待处理案例的特征匹配

该阶段是要将案例的特征与法律领域本体案例库中的本体信息进行比较、匹配，通过本体解析，在法律条例本体案例库中找到与案例的特征相似的相关法条，并提取法条及相关审判推荐。通过与法律案例本体案例库中的特征匹配，找到与案例相关的历史案例，并提取审判结果。

4) 结果反馈

最后，将与案例相关的法条及推荐刑期、相关案例的信息向最终的结果展示层进行反馈。

1.6 章节安排

本书的研究章节安排分为四部分。

第一部分（第 1 章、第 2 章）：绪论和国内外研究综述。第 1 章主要论述了本书的研究背景、研究意义、研究目标和研究方法。第 2 章对本书相关领域的研究现状和有关文献进行了综述。

第二部分（第 3 章）：基于本体的刑事案例库构建研究。它包括基于本体模型、对半结构化刑事案件信息结构化表示、基于本体和案例文档集对刑事案例信息进行抽取，并将抽取的信息进行结构化存储，形成支持决策的案例库，目的是实现决策信息的可推理，解决复杂决策的信息不完备性问题。

第三部分（第4章）：基于遗传算法的刑事案例属性优化与检索策略研究。它包括案例属性的优化方法论证与案例检索的方法论证，目的是为案例推理提供可靠的相似案例库，实现对决策推理机制的支撑。

第四部分（第5章、第6章）：刑事案件推理集成决策模型研究和刑事案件量型决策辅助系统原型设计。它包括两种推理方法的集成决策模型构建、关键模型的训练与使用和决策模型的实验验证，依据耦合理论，将特征选择、决策树、神经网络等数据挖掘算法融合到规则推理与案例推理之中，发挥各自优势，支持量刑决策，并根据研究成果建立辅助决策系统原型，实现管理决策理论与信息技术、领域知识相结合的可视化。

本书的基本研究框架如图 1-4 所示。

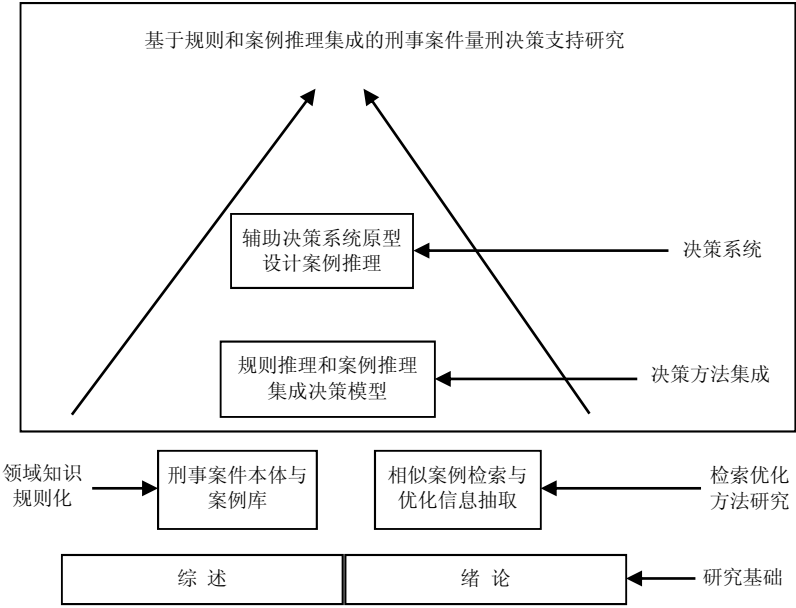


图 1-4 本书章节结构框架

1.7 研究的创新之处

本书的创新点如下：

第一，基于本体理论，阐释了领域知识组织、信息抽取和案例存储三个递进环节组成的半结构化文本规则化、结构化机制，有效解决了半结构化数据支持决策基础问题。这一机制在刑事审判领域的创新性应用，大幅拓展了支持决策的信息范围，解决了复杂决策的信息不完备性问题。这是目前在地区法院解法律文本结构化与数据可挖掘及应用方面具有拓宽性的研究。

第二，针对刑事案例多属性、多要素特点，通过比较提出和论证了对于多属性、多案例、多要素背景下，基于遗传算法的刑事案例属性优化方法，运用 GA-KNN 方法的结合，实现了多案例、多属性、多要素司法案例推理中的案例检索的可靠性与高精度，解决了支持刑事案件量刑决策可靠案例的关键技术问题，提高了司法案例检索的精度和可靠性，拓展了多类型、多属性案例的检索与属性优化方法。

第三，构建了支持刑事案件量刑决策的规则与案例推理集成决策推理模型，实现了决策树模型支持的刑罚规则挖掘与神经网络模型支持的量刑决策推理，创建了管理决策理论、数据挖掘方法和司法审判领域规则相融合的便捷、高效的支持决策模式，并通过系统原型设计对此进行了验证，拓展了领域知识挖掘和利用机制。

三个创新点相辅相成。创新点一解决决策基础问题，它是决策过程中的一个前提或基础，决策时要对案例操作，而案例的表现形式是半结构化的数据，为此只有将半结构化的数据规则化形成可挖掘的数据库，才能使依赖半结构化

数据的决策成为可能。创新点二是在创新点一构建案例库的基础上解决决策关键技术问题。决策推理的前提是必须有适合决策推理的案例库，只有满足快速检索和适合的案例属性才能使推理成为可能。创新点三以相似案例库为基础解决决策推理的理论和方法问题。基于相似案例，对案件进行分类模型和刑期推理模型训练，形成案例属性与刑罚类型对应的规则集，形成支持推理的神经网络模型。同时，创新点三将本书的成果通过辅助系统原型设计进行实例化。三者是递进式关系，三者共同构成支持刑事案件量刑决策的必经途径。

国内外研究综述

实现基于规则推理和案例推理的量刑决策模型，就需要考虑如何进行案件相关信息的存储和规则化、如何进行推理找到“刑”、“期”与案情信息的关联、如何构建决策支持平台的问题。本章将结合上述问题对国内外研究现状进行综述，力图通过文献的整理，寻找解决问题的思路和研究方法。

2.1 法律知识的表示与抽取

近十余年来，语义技术、人工智能和法律研究领域取得了显著的进展。本体模型应用于多个领域的知识表示和存储，其在语义关联方面相比于关系数据库更加灵活，使非结构化文本的处理和机器学习成为可能。

2.1.1 法律知识的表示

1. 本体模型与法律本体

1) 本体模型

本体模型是一个特定领域的概念模型。Gruber^[9]将本体定义为一个概念化的规范,更具体地,它是作为描述存在于某个体或群体中的概念和关系。本体模型在多领域广泛应用,由于目的各不相同,所以目前还未有一个确定的定义和表示方法。Mommers 等^[10]认为,计算机科学和人工智能中设计的“本体”与哲学背景下的本体术语有显著区别。在人工智能中,本体模型主要用于提供知识表示的基本框架:实体和关系,并提供给法官一种表示知识本体所覆盖的领域的方式。Wyner^[11]提出,本体模型是一个明确、正式、一般的概念化特性规范和给定区域目标间的关系。对于一个本体,可以给定对象类及适用于对象类的属性、属性特征和限制条件;并指定各种本体对象之间的关系,和类之间的继承关系,同时进一步定义其特定属性。Valente 和 Breuker^[12]认为本体模型能够建立法律理论和人工智能之间的缺失环节。如果将本体模型理解为论域的形式化描述^[13],那么法律本体可以认为是法律论域的形式化描述。

2) 法律资源信息结构

法律文本的主要特征可以总结为:结构良好,即较易于设计逻辑本体,但在通常情况下,大型文本语料库中法律知识是以复杂的方式分散分布的。从本体建立的角度来说,法律资源是一个复杂的多层信息结构,由底层向上,包括以下5个方面。

(1) 文本。文档中得到具有法律效力的部分。

(2) 文本的结构。文档中文本组织结构的部分。

(3) 元数据。元数据可以包括文档描述元数据（如关键字）、工作流程（如审判的程序步骤）、生命周期文档（如该文件在那一段时间的历史文档）、文档标识元数据。

(4) 本体建模。在现实中，法律文档扮演的角色（如判决的司法概念），关于此种现实角色的任何信息或文本中的任何概念都需要建模。

(5) 法律知识表示。在法律意义下，对文本含义部分的解释和建模。

法律本体的作用。本体在人工智能和法律领域中地位的不断提升，既可以看作是计算机科学和人工智能发展的必然趋势，也可以认为是建立可维护、可扩展的法律领域结构模型所必需的。Deborah McGuinness^[14]指出本体工程领域已经成熟，对法律本体语义技术应用的研究日趋增多，促使本体工程领域积累了丰富的经验。法律本体是多样化法律应用的概念模型，如信息检索、交互服务和推理，而法律本体发展的需求及法律知识管理应用的多样性都正在增长^[15]。法律知识管理应用的多样性包括文件检索、数据和文本挖掘、计算机辅助法律起草、合理化的法律领域（如编纂）、法律推理模型、多代理模拟、决策支持系统开发环境或资源管理等。

法律本体的作用可以总结为如下三点：

有助于编写程序用于推理并管理概念类型间的关系和差别^[16]；促进在案例库中或分散在互联网中其他资源的知识和信息的交流和使用^[17]；自然语言解释生成，进行决策支持。

2. 法律本体建立理论

1) 面向语义的理论

人工智能和法律中应用的若干逻辑语言通常使用语法和语义表述特定的法律概念（法律类）来做出本体假设。基于此，道义逻辑使用上述基本假设区分不同类型的道义。例如，McCart 提出的面向语义的法律论述语言，它基于语义和本体论框架，区分了空间、时间、质量、行动、权限、责任、因果、目的、意图、知识和信念。

对于非结构化的法律文本，在法律领域进行面向语义的本体建立较多地使用 WordNet。WordNet 是概念之间的关系框架，概念是由一组同义词（即同义词集合）与相应的含义解释共同表征的。这些同义词集合通过一组固定的关系相互链接^[18]，包括因果关系和其他相关性。可以将 WordNet 理解为一本自动化的语义字典，它使用计算机程序对某个词语的意义进行定义和解释。

JurWordNet^[19]是对通用 WordNet 的扩展，即将通用资源与专业资源的术语集相链接，包含法律术语集，其同义词集之间的关系不仅包含用于分类的垂直关联，还包含水平关联，用以对语义等价做出限制。LOIS WordNet^[20]由欧盟资助研究，是支持六种欧洲语言的法律 WordNet，部分基于 JurWordNet 且包含大约 5 000 个法律概念。上述法律领域 WordNet 将非专业的搜索术语与专业法律语言相链接，便于在法律文件中进行搜索。综上所述，WordNet 是一个广义的语义框架，它使法律词汇网络与通用的词汇本体相关联。

2) 面向认识的理论

法律通常被认为是一种人类构造的现象，从认识的观点看：它主要是人与人之间的协议，且是无形的。Valente A 等基于面向认识的理论建立功能本体。

功能本体将知识分为 6 种不同的基本类型：规范性知识、世界知识、法律元知识、责任知识、活性知识和创造性知识。从信息科学的角度出发，该法律本体模型遵循由 Gruber 给出的本体定义和相关法律理论。

3) 面向本体的理论

面向本体的理论与现实密切相关。早期面向本体理论是 Van Kralingen^[21]提出的“基于本体的框架”。该理论将法律实体分为三种主要类型：行为、规范和概念。其中每一种实体类型都由框架表现：行为框架包括行为标识符、公布、范围、代理、行为类型、手段、方式、时间、空间、情况、原因、目的、意向性和最终状态；规范框架包含规范标识符、传播、应用的范围、条件、主体、法律形态和行为标识符；概念框架包括概念、概念类型、优先级、颁布、范围、条件和实例。Verheij B 和 Hage J^[22]提出包含三个主要元素的模型：事件状态、事件和规则。事件状态通过描述性的句子表述，事件带来的变化通过当前事务的状态体现，规则直接表达在事务之中。因此，在事物状态之间有两种不同的关系：第一种，从一种状态过渡到另一种状态是由一个事件引起的；第二种，从一种状态过渡到另一种是规则的结果。该模型受法律制度理论启发，通过调节当时事务和事件的关系，以及事物状态之间的偶然性关系，对法律现象提出了一种非常抽象的建模方法。

上述三种理论各有侧重，面向语义的建立理论通过对自然语言的处理和分析建立法律术语词典形式的语义本体；面向认识的理论从法律概念本质观点的角度出发，对不同类型的知识分类以构造本体；面向本体的理论从框架和事务关联的角度对本体进行较为细化、关系化的建立。将上述三种理论加以整合可以更好地完成法律本体的建立，如 Mommers L^[23]提出“基于知识的法律本体”，通过区分本体论的认知角色和地位层次，调整了在单一的本体中的认知和本体

论的观点。本体中包括一些面向语义的类别,但该方法的主要关注点是适应现有的法律系统及法律知识。另外,文本挖掘也成为法律本体建立和信息处理过程中的关键点,文章^{[24],[25]}中详细介绍了相关研究。

3. 通用本体与领域本体

法律本体通常分为通用本体和领域本体。通用本体通常是在法律专家的知识之上建立的,包括法律理论基础概念的形式化。领域本体包含关于规范知识的明确假设、关于现实法律的理想观点、由实体组成的知识及社会现实等,通过施加约束使规范生效。

通用本体主要包括:最早的法律领域深层次概念化模型 LLD^[26]、法律功能本体 FO-Law^[27]、基于框架的概念性法律本体 FBO^[28]、基于知识的法律本体 KBO、以 FO-Law 为基础发展的 LRI-core^[29]、实现法律知识系统知识交互的法律核心本体 LKIF-core^[30]、提供多语言和异构信息源的 JurWordNet 等。其中,FO-Law 法律功能本体模型按照功能的不同将法律知识分成许多相互联系的类及子类,从法律在社会中所扮演的角色和功能的角度出发描述了法律知识,具有推理及文集解决的功能;JurWordNet 具有组织、构造信息的功能;LLD、框架本体和基于知识的法律本体是用普通语言来表达法律知识;FO-Law 探索提供法律问题解决方案;而 LRI-core 和 LKIF-core 支持了领域本体知识的获取。这些法律本体能够更好地表达法律知识,对法律信息进行分类,能够实现共享法律知识的重用。

在通用本体的研究基础上,研究者开始更加注重对具体领域的知识表达,产生了越来越多对法律领域本体的研究。Peters W 等^[31]提出了 LOIS 工程的法律知识结构,通过将词典中的语义和具体法律领域中的语义联合的方法来表示法律概念,提供相应的法律知识,以优化法律信息检索策略。Asaro C 等^[32]认

为领域本体的概念化可以通过元数据的方式管理文件、用 XML 对刑法进行语义标注,为法官鉴定或建议犯罪假设提供决策支持。Lame G^[33]通过收集的法律文件建立法国法律本体,运用本体对法官的查询需求进行扩展,缩小搜索范围,提高搜索的精准程度。

综上所述,法律通用法律本体面向法律的通用概念进行建立,领域本体则是针对某一具体的法律领域或引用领域建立。随着程序设计中模块化的设计思想的成熟和广泛应用,以及本体规模的逐渐增大,本体结构也在向结构化和层级化的方向发展。例如在专利研究中,Giereth M 等^[34]在实现语义内容生成过程中,使用多个本体描述专利信息,包括通用顶层本体、用于专利信息模型描述的专利顶层本体(由专利元数据本体、专利结构本体、专利分类本体构成),以及用于语义标注的语言学本体、语义本体和法官本体。在未来的研究中,将法律通用本体与领域本体、语义本体等相整合的方式和接口设计可能成为法律本体设计的研究方向之一。

4. 法律本体的建立方法

自动化的法律本体建立与文本挖掘、模式识别联系紧密,以下研究文献均使用自然语言处理的相关方法建立法律法规本体。Sylvie Despres 等^[35]使用半自动的 TERMINAE 方法和法律核心本体 CLO,对欧盟指令中的法律指令集进行微本体建立,并实现本体融合。首先,在微本体建立的过程中,对指令集文本进行分词,形成语料库;运用 Syntex 进行词标注,并进行概念之间关系模式的探索。然后,对处理好的语料进行本体建立:将词语转化为概念,关系转化为角色,识别结构属性和功能属性。在此基础上进行巩固建模,实现概念的一般化、特殊化,以及对概念聚类属性的重组,并且用 CLO 进行校验。Cestnik B 等^[36]人以政府公开招投标文档为例,提出一种通过建立本体以实现清晰的文档结构和

内容展示的方法,提升公众对法律文档的理解程度。文中指出,对文档的处理要经过多次如下步骤的循环:首先,对文档进行段落分解;其次,通过词形还原和停用词去除来获得段落中领域相关的重要概念;第三,运用 OntoGen 工具在专家的辅助下建立顶层本体概念,并形成关联;第四,在已有的分段结果和本体概念的基础上形成对文档内容结构的可视化;第五,识别分配不合适的异常段落并进行修正处理。Lame G^[37]提出基于 NLP (Natural Language Programing, NLP) 技术的从法律文档中识别法律概念和语义关联的方法。首先,进行法律概念的识别,以法国法典为语料库,结合 Syntax 工具进行文本句法分析,并运用句法规则对目标文本进行概念抽取,再结合统计指标对抽取的概念进行筛选,确定法律概念与非法律概念,去除无实意词等。其次,通过对句法分析、关联关系分析、统计分析和模式匹配来识别概念之间的关系,最终建立本体。

相比结构化程度较高和语义描述特征性较好的法律法规而言,对法律案例的本体建立更为复杂,常与案例的推理相关。Adam Wyner 描述了两用于案例检索、支持决策的基于法律案例的推理系统。对案例的表征和抽象主要采用判方和因素层次结构相关联的方法,通过对案例进行因素标识,在检索中进行因素相似度比对,并依据因素偏好确定判方来辅助决策。Kingston J^[38]以建立金融诈骗案推理所涉及的本体为例,从法官需求的角度出发,以 wigmore 图建立诈骗案模型和相关法律模型。从案例的角度出发,模型由上至下分为假设层、法律层和证据层,下层内容对上层起支持作用;从相关法律角度出发,分为议题层、子议题层、条件层、证据层,并将多种通用法律本体与金融本体相融合,实现对金融诈骗相关案件的决策支持。

2.1.2 法律知识的抽取

1. 文本挖掘

以文本挖掘为主要代表的数据挖掘对决策起着重要的辅助作用，特别是针对决策中不确定信息进行综合挖掘，提取具有知识性、规律性的内容，是当今挖掘技术和决策研究的热点。文本挖掘就是从非结构化、半结构化的文本中发现潜在的概念及概念间的相互关系。大多数学者集中在中文分词、文本特征提取、分类、聚类算法上，系统地将这些方法融合在一起应用于实际领域的研究还比较欠缺。李子叶等^[39]基于模糊集贴近度，提出了文本信息分类器；王丽珍^[40]通过面向属性归纳，提出了基于语义贴近度的抽象归纳法，它通过将低层属性归纳为高层概念，进而有效地提高数据挖掘效率，发现不同层次上的知识；李蕊等^[41]对基于语义相关和概念相关的文本自动分类研究；程少川等^[42]针对知识管理中的信息组织难题，在非结构化数据（包括文本）的挖掘中建立本体系统进行了研究，对本体系统方法的概念、构成和应用进行了示例。与此类似，王娜、李云松^[43]提出了利用概念格来抽取隐含在文本中的概念关系，将文本挖掘中文档与关键词之间的关系利用概念格结构呈现出来。在数据挖掘中，针对连续特征离散化也是一个重要研究内容，汪庆等^[44]对此进行了综述。在文本挖掘中，知识获取的不确定性主要来源于有限的分辨能力及对于数据描述的不确定性，马志锋等^[45]对规则获取进行了研究。陆光辉、肖人彬^[46]从大型数据库中获取有用的知识出发，提出了基于粗集理论和神经网络的集成化数据挖掘方法。徐亚娟^[47]提出一种基于案例相似性计算机的文本信息挖掘模型，从文本预处理、案件特征选择、案件分类挖掘和应用系统设计等方面对公安犯罪案件文本挖掘关键技术进行了研究，在公安业务信息挖掘中得到初步应用。

利用文本挖掘对大规模真实文本进行处理的关键技术，一是信息抽取技术，

要求算法能有效处理大规模文本并能够从文本中抽取有用信息；二是针对行业应用，以建立影响决策的行业本体，利用本体挖掘技术，进行有效的数据挖掘；三是对挖掘结果的可视化。通过文本挖掘，进一步了解和掌握知识的存在形式^[48]。

2. 文本信息抽取

文本信息抽取是文本挖掘和数据挖掘的前提，更是数据转换和装载到数据仓库的必经阶段。Sharjeel Imtiaz 和 Azmat Hussain^[49]提出了一个基于多 Agent 的基础框架，集成了文本特征挖掘和关系数据挖掘。Dozier C 和 Jackson P^[50]认为专家型证人在解决法律官司中起到重要的作用，提出算法解决如何找到合适的专家证人并评估、比较他们的能力。Kenneth E. Hild II^[51]提出基于信息理论特征抽取，用二阶统计量来比较线性关系间的随机变量；用信息理论方法来比较非线性关系间的随机变量，如特征向量和类标签，评价标准是交互信息。Margherita Berardi 等^[52]以生物医学领域为研究背景，以 HmtDB 数据更新抽取为主要实验内容，搭建了一个信息抽取的框架。Lipika Dey 等^[53]给出了通过对文档的语义和语言分析识别半结构化信息组成成分的方法论。一个多角度的校对机制，实体的角色由实体在句子中出现的关系所决定，关系的识别是通过语义和语言分析得到的。XiaoLing Wang 等^[54]使用 word sense 消除歧义法来得到备选关键短语的 word sense，并且通过词汇归并、特征计算和评估来考虑备选关键短语的 word sense 的语义相关性。这种关键短语抽取方法与当前传统的方法不同，使用的是 senses（词义）集而不是 terms（词汇）集。Hiep Phuc Luong 等^[55]建立了一个基于语料库的本体学习的框架，包括自动支持检索文档的任务，为充实本体过滤和抽取相关信息，使用了一个焦点爬行器采集文档和信息。Tao Xie 等^[56]研究和建立了一个面向复杂结构的精确 Web 文本信息抽取过程化模型和规则系统。Shady Shehata 等^[57]建立了一个基于概念的挖掘模型，它能够

在句子、文章和语料库的层面上分析词语。Christopher Baker^[58]提出抽取变异并且找到变异相应蛋白质序列的变化主要包括两个步骤：第一步，从文档中辨别出变异实体和蛋白质实体；第二步，辨别目标蛋白质的序列并正确定位变异发生的位置，给出了对应的算法。Christina Feilmayr^[59]建立了将数据挖掘与信息抽取相结合的信息抽取系统的方法论，包括语法结构、词性标注、消除歧义和语义等过程的预处理、基于规则的知识抽取、基于特征选择的知识推理求精、基于领域本体理论的背景知识处理等。Ning Zhong 等^[60]给出了一个有效的模式发现技术（算法），主要的特点是根据模式中词的属性而不是文章中词的属性来确定的权重可以提高词语权重的准确率。Luis Tari 等提出了一个新的信息抽取方法，即在预处理阶段将非结构化的数据以“树”的形式进行存储，这种“树”的结构将能最大限度地保留原始语料中的语义关系。

Benjamin Adrian 等^[61]将领域本体整合到信息抽取任务中，提供一种基于管道架构、基于本体的信息抽取方法。运用 RDF/S 来描述术语及领域本体实例，识别领域本体词汇要素：类层次、属性层次、实例、特征和推断，并基于这些特征进行特征识别、实体识别、事实识别、模板群识别。iDocument 提供一种基于 RDF 查询语言 SPARQL 语义的常用的标准 IE 模板接口，通过 SPARQL 描述本体的 RDFS 模式来表达抽取模板，最终生成一个从文档中抽取出来的服从 SPARQL 模板的包含 RDF 三元关系的 RDF 图，供法官或专家进行进一步分析。

M. Saravanan 等^[62]针对法律领域的案例文本，提出一种 SMILE 机器学习（Machine Learning, ML）方法，这种方法受 ML 和 IR 在文本分类方面的成功研究的启发，不是采用统计模型，而是采用一种特征 ML 的方法。这种方法更适合处理短案例文本的收集，同时有助于对背景知识进行整合。SMILE 的整个流程主要分为样本学习和新案例标注两部分。在样本学习阶段，输入的内容是已经经过特征对应标记的样本集，有标记的句子表示一个特征集。运用特征学

习算法来进行规则识别，并将规则提供给分类器。分类器用于给新案例文本进行标记。首先，识别句子模式，抽取姓名，并进行角色对应；其次，在多重特征中发现行为，运用人工自然语言处理工具进行命题模式的发现，捕获词之间有意义的句法关系；再次，由于法律案例中有明显的对抗性，否定词表达的否定意义会关系到因素的选择和表征，因此在算法中加入“no”和“not”来辅助因素的选择，增加算法的准确性；最后，采用树结构对识别的规则中的角色进行深度匹配，从语义、语法结构等最终确定因素标记。经过最终标记的文本主要用于案例检索的查询匹配。

3. 文本语义标注

语义标注是对于文本信息，实现人机交互的基础。对于计算机来说，大量的文本资源具备“可读性”，但没有“可理解性”。这正是语义标注的工作。语义标注是自然语言处理 NLP 的一个重要应用，其任务是将文本中涉及的实体（作者实体、术语实体、机构实体等）与其对应的本体实例（语义描述）相关联，自动对大量文本进行标注，指出资源在语义层面的特征。通过在文本中加入标记或标签，将文本中的实体与其对应的本体实例，即语义描述相关联^[63]。

语义标注系统所采用的方法，从人工参与程度来分，主要分为：手工标注和（半）自动标注两种。

1) 手工标注方法

手工标注方法是采用事先定义好的本体对文本中存在的实体进行的标注。标注人员需要对本体结构及待解析的文本结构、元数据元素等内容进行全面把握，同时能够运用 RDF 或 HTML 语言标记语义数据。

手工标注方法的一般流程为：打开待标注文档并指定相应的本体；抽取文

档中的关键词，将关键词与本体中相应的概念相关联，以建立新的标注实例；在本体的指导下，为新建的实例添加相应属性；标注人员为实例添加各属性值；将标注结果以 RDF 或 XML 的形式保存到原文档中或统一保存到标注库中。

目前支持手工标注方法的系统主要有：SHOE、Ontobroker、OntoMat、Annotea 等。这种方法的标注精度虽然很高，几乎为 100%，但仅适用于处理小范围文本的标注，对于越来越多的大规模海量数据，这种标注方法显然难以胜任。

2) (半) 自动标注方法

从手工标注向自动标注过渡的过程中，还有一种半自动语义标注方法。通过分析文本识别文本中的实体，然后将分析得到的实体与其对应的语义描述相关联，同时在人工的介入下，将一些含有歧义的词条进行分类，因此称为半自动标注方法。随着科学的进步和技术的不断积累，目前已经有一些全自动的语义标注工具了。现有的自动语义标注的方法主要有基于规则的语义标注、基于分类模型的语义标注、基于序列模型的语义标注、Deep Web 标注等。

(1) 基于规则的语义标注。基于规则的语义标注包括：基于手工规则的标注和基于规则学习的标注。基于手工规则的标注是由法官定义标注规则，系统根据标注规则对文档进行标注。手工规则多通过定义“Wrapper”^[64]的方法实现，运用标注信息的上下文识别文档的语义信息。Kim^[65]平台中的语义标注模块利用 GATE 实现了自然语言的处理，如分词、词性标注、命名实体识别、规则匹配和指代识别等，并将其进行集成。Popov 等通过对 GATE 已有规则的扩展实现了对新闻数据的自动语义标注。Alani H 等^[66]利用自然语言分析工具进行语法和依存关系的分析，结合 GATE 和 WordNet 识别文档中的命名实体，并通过词义分析实现概念关系的标注。Leonard T 等^[67]针对 XML 和 HTML 文档

通过记忆法官的操作流程实现语义标注。Kogut P 等^[68]提出 AeroText 信息抽取模式,通过预定义的规则将文档中的专有名词及名词间关系映射到本体中相应概念和属性上,但其支持的本体语言有特殊要求。手工定义规则的方法可以提高标注的效率,但是需要在标注前人工地对文档及数据的结构进行详细分析,规则的适用范围有限,很难在所有数据上都取得很好的标注效果。

基于规则学习的标注方法通过给定的训练数据自动学习标注规则,并对规则进行不断的修正,并用最终的规则集进行新文档的标注。Ciravegna F^[69]提出了规则学习算法 LP2,并开发了自动语义标注模块 Amilcare。首先利用 GATE 对文本进行预处理,然后运用 Amilcare 从训练集中学习识别标注信息的起始边界和终止边界的标注规则,建立初始规则集,接着运用一个覆盖度算法对规则进行归纳,并进行不断迭代。这种标注方法已经被用到多个语义标注系统中,如 MnM^[70]、Melita^[71]等。Benjamins R 等^[72]针对 Web 页面,使用包装器学习技术实现自动标注,除此之外,还支持图像和多媒体数据的标注。WebKB^[73]利用 SRV 算法^[74]学习标注规则。这种标注方法适用于具有一定模板结构的文档,但是由于规则学习方法主要还是采用随机归纳的方法,学习效率低,质量也不高。

(2) 基于分类模型的语义标注。这种标注方法将标注问题看作分类问题,利用统计学习的方法,如支持向量机^[75]、感知器模型 Preceptron^[76]和贝叶斯模型等,来进行标注模型的学习。利用分类模型,可以通过对候选标注对象的特征提取,学习标注信息本身的特征分布,用训练得到的特征模型判别候选对象的归属,也可以利用特征模型来判断候选对象的起始和结束边界,将这之间的文本作为标注结果^[77]。分类模型具有比规则归纳更强的学习能力,可以表示更加丰富的特征分布。

(3) 基于序列模型的语义标注。序列模型利用标注信息之间的依赖关系,对输入序列统一进行标注。在文本标注中,序列可以是文档中所有文本行的序列,也可以是单词的序列,标注的任务是为这个序列中的所有单元给出一个全局最优的标注结果。Reeve L 等人提出了用马尔科夫模型进行语义标注的方法^[78],其他常用的序列模型包括:隐马尔科夫模型、最大熵模型、最大熵马尔科夫模型及条件随机场模型等。序列模型增加了对标注信息之间的依赖性的考量,提高了标注精度,但其主要描述线性特征,对于非线性依赖的关系标注效果不一定很好。

(4) 其他语义标注方法。其他语义标注方法有 DeepWeb 标注、依据语句成分依存关系进行标注、基于语义排歧的标注及基于无监督学习的标注等。Deep Web 标注是在深网^[79]的概念中衍生出的标注方法。袁柳等^[80]将领域本体作为 Web 数据库遵循的全局模式引入到查询结果语义标注过程中,通过对查询接口及查询特征的详细分析,采用查询条件重置的策略来确定查询结果数据的语义标记。Handschuh S 等^[81]通过标注数据库模式和本体之间的映射关系来实现对法官请求的动态反馈。依据语句成分依存关系进行标注的方法是通过构造自然语句中的词语与本体的对应和词语之间依存关系与本体概念之间关系的对应来实现从依存关系到 RDF 三元组的转换,实践系统如 Jianming Li^[82]构建的 ALPHA 系统。基于语义排歧的标注方法主要是借助统计的方法,对词出现的上下文特征进行统计和识别,进行意义排歧,如 Dill S 等。无监督学习的标注方法和语义排歧的标注方法在上文所提到的一些方法中都有涉及。无监督学习的标注方法通常是基于规则学习,结合自然语言处理技术对抽取规则进行不断的优化,是对预定义规则标注方法的改进,实践系统如 Philipp Cimian 等^[83]建立的 CPANKOW 系统。在实际应用中,通常是将各种方法综合运用,实现优势互补,能够更好、更准确地抽取所需信息。

总结起来,目前语义标注工具所采用的方法主要有三种^[84]:一种是基于模式的匹配,一种是基于机器的学习,还有一种是将前两种方法混合在一起,又称混合模式,充分运用各自的优点互补对方的劣势。具体的语义标注方法如图 2-1 所示。

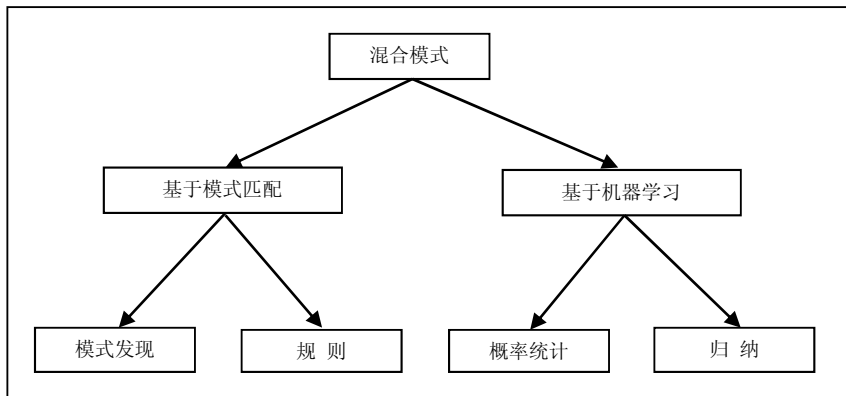


图 2-1 语义标注采用的方法

4. 基于本体的语义标注系统

将整个本体标注过程作为一个系统,则目标文档和本体即为系统输入,系统输出为本体实例或个体的属性值,而系统的关键部分即本体标注部分,实为对文本的自动分类,可以有很多算法,如基于 WordNet 的语义相似度算法、基于正则表达式的信息抽取算法。如图 2-2 所示,戴维民等^[85]给出了一种基于特征向量分析的本体自动标注模型。

Atanas Kiryakov 等^[86]结合 KIM 平台阐述了自动语义标注的整体流程,并指出语义标注的三项基础:定义实体的类或本体、识别或链接到语义描述的实体标识器、包含实体描述的知识库。

Hammond B 等^[87]提出的 SCORE Enhancement Engine (SEE) 是一个利用

分类模型自动标注文档类型的系统，它利用分类学习模型，判别文档的语义类别(类似文本自动分类)，然后利用统计学习方法识别 Web 页面中的命名实体，最后通过对识别结果和 Web 页面的分析实现实体关系的标注。

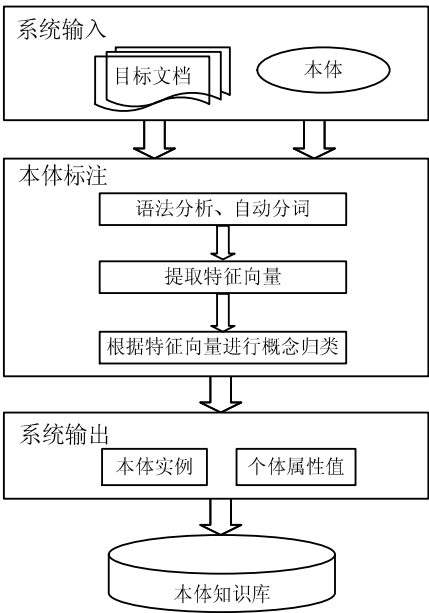


图 2-2 本体自动标注模型

荆涛等通过结合通用标注和本体标注方法，实现语义标注，如图 2-3 所示。首先在数据准备阶段，构造领域本体及领域词汇表；在识别阶段通过构造基于词汇建立的本体类型标注器和基于正则表达式规则的通用类型标注器，以句子为单位，对句中的词汇标注对应类型；在组合阶段，通过分析句中词汇间的依存关系构造语法关系三元组，并依据类型标注器产生的标注结果和领域本体知识，生成并验证知识三元组，即表示为 RDF 陈述的形式。

高琦^[88]提出了一种基于弱监督的 Bootstrapping 和基于规则的本体标注模型，结合大规模的无标注语料来改善文本分类及标注性能。整个标注模型是一

个不断迭代的过程，每次迭代分为本体解析、文本分类及信息标注抽取三个部分，通过不断迭代对本体进行丰富。

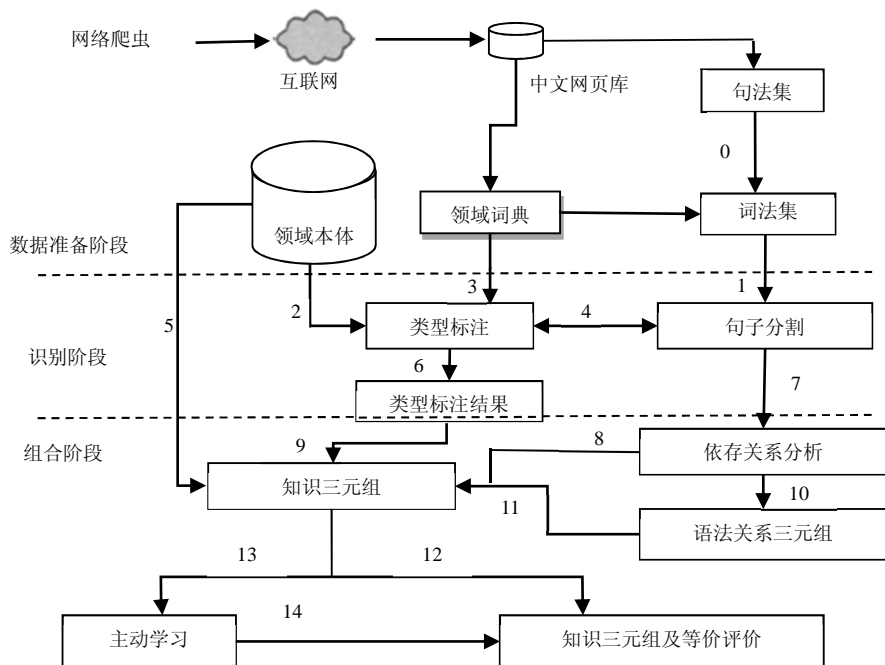


图 2-3 中文网页标注方法

Sushain Pandit^[89]提出了基于解析树和依赖树的文本信息抽取方法。首先结合句法分析，逐句处理文本，对每个句子生成解析树和依赖树，运用抽取规则识别和抽取潜在的关系和实体；然后用领域本体模型和这些潜在者进行匹配，同时可以对模型进行新概念和关系的添加；最后用 RDF 语法表示这些潜在的实体和关系。

Theresia Gschwandtner 等^[90]针对医学文档建立了一个交互式的半自动标注系统 MMTx，以 UMLS 原词表为概念抽取模板，并将抽取结果导入到 MapFace editor 中，从概念层和句法层供法官对抽取结果进行修正。戚欣等^[91]在基于本

体案例库进行自动语义标注的过程中,为识别命名实体,设计了语义词典的逻辑结构,建立了词语到语义对象的映射,同时提出了基于最短语义关联路径的语义消歧方法和基于 N-gram 的语义消歧方法。常平梅^[92]提出多本体支持的语义标注模型,它包括本体集成模块、信息抽取标注模块、信息检索模块。本体集成模块对关联的多个本体进行集成,然后解析集成后的本体,根据本体知识生成 JAPE 规则;在 JAPE 规则的指导下,信息抽取标注模块借助基于本体的信息抽取标注抽取概念、实例和关系,并按文档综合相关度算法,通过计算文档相关度,整合标注信息并将其存入标注数据库。

2.2 法律案例的属性优化与案例检索方法

在法律研究领域,得出判决结果的逻辑推理过程一直是学者研究的重点和难点。从推理过程所依靠的“事实”的角度分类,可以分为基于规则的推理(Role-Based Reasoning,简称 RBR)和基于案例的推理(Case-Based Reasoning,简称 CBR)。其中基于案例的推理,即由目标案例的某些因素或特征而检索得到案例库中的基础案例,并由基础案例指导目标案例求解的推理策略。要快速检索到基础案例,就需要对案例属性进行优化,并选择恰当的技术实现快速检索。

2.2.1 案例属性优化

在针对案例的数据挖掘中,案例的属性选择比较复杂,必须要根据挖掘的目的、所在的知识领域和需求对象,进行算法描述和挖掘任务的规划。为了确

保案例推理的效率,需要对案例的属性进行处理,也就是属性优化与约简,确保检索效率。常用的案例属性优化方法有模糊综合评判法、模糊优选法、层次分析法(AHP法)、粗糙集法、定性指标综合方案法、加权相对偏差距离最小法等。目前,运用比较广泛的就是AHP法和粗糙集法。AHP法给决策者解决那些不容易定量描述的决策问题带来了极大的方便;粗糙集法易于知识系统的表达,缺点是可能在离散化后造成信息丢失,特别是不能处理所有属性。

属性权重是属性重要程度的一种主客观反映,它在属性约简中是非常重要的因素,但在当前的案例检索中,权重选择、属性约简和案例检索都是孤立完成的,或者是层梯式使用,没有完全结合起来考虑,针对将案例、属性、权重集为一体的研究越来越引起人们的重视。1999年,Kuncheva and Jain提出了基于GA的特征选择和案例选择的两维度同时优化方法;2003年,Rozsypal Kubat将编码方法和适应函数等进行了优化,并通过实验证实了这一方法的优越性,两维度同时优化明显地减少了案例的搜索空间;2006年,Hyunchul Ahn和Kyoung-jae Kim把这种算法用在了客户分类系统中,从而在实际应用环境中对此进行了检验。但是,现在有学者发现这种基于遗传算法的优化虽然提高了命中率,但是没有考虑到属性权重对于属性约简的重要性,在实践中,可能一些很重要的权重在优化的过程中就被约简了,有一些不重要的权重又没有约去。为了解决这种问题和考虑到特定领域自身属性比较多、在属性权重上的差异也比较大的特点,我国学者李海芳等^[93]提出了AHP层次分析法与遗传算法GA相结合,实现案例库的缩简、属性的约简和权值确定这三方面进行综合检索优化的方法,建立了一种检索模型,并将这种模型运用到基于旅游的多策略挖掘系统进行实验,结果表明在案例检索的命中率上有明显提高。

1. AHP 方法的属性优化

AHP 法^[94]把复杂问题的组成因素按支配关系分组形成有序的层次结构,最上层为目标层,中间为准则层,最底层为方案层,采用两两比较的方法确定层次中各因素的重要性。其求解过程主要包括三个部分:建立层次关系、构造判断矩阵和层次单排序及一致性检验。常用的相对权重的计算方法有和法、根法、特征根法、对数据最小二乘法等,最适合计算机运算的也是最常用的是前两者。

2. GA 方法属性优化

案例属性优化的目的是获得更加具有代表性的案例空间和案例属性,GA 的同时优化^[95]方法基本过程如下:

第一,根据遗传算法的基本步骤,根据可能的所有解,确定遗传算法的编码方式。常用的有浮点式和二进制两种。

第二,利用随机函数产生初始种群,利用初始种群确定适应度函数。遗传算法推理的目的就是通过进化得到使适应度函数达到最大的个体。

第三,确定相似度量算法,以支持 CBR 推理过程中的检索需要。在测试集案例库中寻找相似案例支持推理时,最常用的是 KNN 邻近算法。

第四,重复第三步,直到满足遗传算法结束条件。当满足条件时,也就找到了满意的属性和案例特征。

在模糊综合决策中,权重是至关重要的,它反映了各个因素在综合决策过程中各属性所占有的地位或所起的作用,直接影响到综合决策的结果。设因素集为 $U = \{u_1, u_2, \dots, u_n\}$ 当前主要有以下几种方法来确定权重:

(1) 专家估测法。将 K 个专家各自独立给出的各因素权重进行平均, 作为该因素的权重。

(2) 加权统计法。当 $K < 30$ 时, 用加权统计方法计算各因素的权重。

(3) 频数统计法。请 K 个 ($K > 30$) 领域专家根据权重分配调研表, 对各因素提出自己认为的最合适的权重, 根据收回的权重分配调研表, 对每个因素 u_i 进行单因素权重统计, 以此得到每一个因素的权重。

(4) 模糊协调决策法。它是分为正问题和逆问题分别进行评判, 最后进行综合的权重获取方法。它虽然是一种近似方法, 但由于运算简便, 又易于操作, 且比较有效, 故不少研究人员乐于使用它。

此外, 还有模型关系方程法, 它在案例属性过多的时候, 计算比较复杂, 不适于做大规模数据的研究。

2.2.2 案例检索方法

案例匹配采用的主要方法有: 最近相邻算法、归纳法、层次检索和基于知识检索等方法。我国学者李锋、魏莹^[96]提出了基于本体给出识别及量化属性间映射关系的方法, 引入语义相似对最近相邻算法; 杨奕飞、瞿元新^[97]构建了基于最近相邻法、归纳法和知识法, 给出面向对象的案例索引机制; Rodriguez M Andrea 等^[98]给出改进的基于本体概念相似度计算方法。案例检索是案例推理系统的中心环节, 检索质量关系着整个系统的质量。李海芳等利用遗传算法 GA 和层次分析法 AHP 相结合, 从案例库属性的约简、权值确定等方面对案例检索进行优化。利用遗传算法在搜索优化上的优势, 使用两维的编码结合权值, 从而形成三维优化, 并利用经验和权值中间表进行权值学习, 从而提高检索命中率。

2.3 法律知识的推理

2.3.1 法律知识的推理逻辑

1. 一阶谓词逻辑

一阶谓词的作用就是用一种特定的模式把知识形式化地表示出来,进而有效、合理地转移并存储到计算机^[99]。一阶谓词逻辑知识表示规范,逻辑性强,推理过程严密,方便计算机扩充案例库,易于实现。

司法三段论可使用一阶谓词表达^[100]:

(1) 对于一切 x , 如果 x 满足了事实构成 T , 那么法律后果 R 就适用于 x 。

(2) s 满足了事实构成 T 。

(3) 法律后果 R 适用于 s 。

法理学中的解释, Tx 是事实构成, Rx 是法律后果, x 可以是公民、法人或其他社会组织。依照法治原则, 即“在法律面前人人平等”, 所以在 Tx 、 Rx 前加全称量词。推理过程为: 某 s 发生了事实, 经查找法律规范发现其符合法律事实 T , 记为具体法律事实 Ts 。由 $\forall x(Tx \rightarrow Rx)$, 根据全称消去规则可得 $Ts \rightarrow Rs$; 由 Ts 和 $Ts \rightarrow Rs$, 根据分离规则 MP 得出具体的法律效果 Rs 。该推理是形式推理, 其有效性不依赖于 Ts 、 Tx 、 Rx , 而仅依赖于逻辑常项 \forall 和 \rightarrow 的性质, 只要 Ts 、 $\forall x(Tx \rightarrow Rx)$ 是真的, 则 Rs 就是一个合法的判决。

2. 描述逻辑与本体

作为概念知识表示和推理的逻辑公式集合, 描述逻辑是基于概念的形式化

表示方法,吸取了 KL-ONE 的主要思想,它是一阶谓词逻辑的可判定子集^[101]。描述逻辑系统有表示概念和关系的子集、Abox 事实断言集、TBox 包含断言集、Tbox 和 Abox 上的推理机制。描述逻辑的表达能力很强,同时具有可判定性,保证推理算法总能停止,并返回正确的结果。描述逻辑的推理主要涉及概念可满足性、包含关系、可满足性、实例检测^[102]。可满足性问题是描述逻辑推理的核心问题,其他许多问题都可转化为可满足性问题。为了能使描述逻辑的可满足性问题实现自动判断, Schmidt-Schaub 和 Smolka^[103]首先建立了基于描述逻辑 ALC 的 Tableau 算法。该算法的基本思路是,把可满足的概念集通过一系列推理规则不断细化、扩展,直到发现冲突。如果直到概念被完全扩展仍然没有不可避免的冲突,就可以得到满足概念集的解,进而证明概念集是可以满足的。

目前大多数的本体模型都是基于一阶逻辑的,如 Ontolingua^[104]、CycL^[105]、LOOM^[106]等。一阶逻辑尽管表达能力很强,但其推理过程相对复杂,不利于本体模型的检验。而描述逻辑作为描述概念和概念层次关系的知识表示语言,虽然没有前者的表达能力强,但推理复杂度是可知的,更适用于需要推理功能的应用领域^[107],特别是描述逻辑的语法容易转换成 XML/RDF 形式,因此基于描述逻辑的本体模型更适合 Web 环境下概念建模和知识共享。

3. 模糊逻辑

模糊逻辑是处理部分真实概念的布尔逻辑扩展。经典逻辑坚持所有事物(陈述)都可以用二元项(0 或 1、黑或白、是或否)来表达,而模糊逻辑用真实度替代了布尔真值。这些陈述表示实际上接近于日常人们的问题和陈述,因为“真实”和结果在多数时候是部分(非二元)的或不精确的。

模糊法学引入模糊推理,替代传统的法律推理,其目的不是将法庭的司法语言直接转换成逻辑形式,而是要提出一种法庭推理的说明模式,并由此指出,在

某种条件下,可以限制任意使用一些不精确的概念^[108]。通过引入模糊算子,凭借人工智能,充分利用法官的能动性,延展人的理性,使司法判决更加趋于精确。

4. 非单调逻辑

单调性和非单调性都是(形式)逻辑系统的特性。判定一个逻辑系统为单调的充分必要条件是,如果一个语句集 S' 是 S 的扩展集,那么,依据 S' 合乎逻辑地得出的结论集 C' 也是依据 S 得出的结论集 C 的一个扩展集;判定一个逻辑系统为非单调的充要条件是,该系统不是单调的。非单调性是相对于单调性而言的,从后承关系的角度来说,所谓单调性是指:如果 A 是 Γ 的后承,那么 A 是 $\Gamma \cup \{B\}$ 的后承,即结论集随着前提集的增长而单调增长。

对法律推理的刻画而言,可能用到非单调逻辑,因为法律推理在某种意义上被明确指定为是可废止的,可废止逻辑的提出者纽特认为人类推理不是也不应当是单调的^[109]。

2.3.2 法律知识的推理方法

1. 神经网络推理

人工神经网络是指模仿生物神经网络的信息处理系统,属于人工智能的计算智能研究领域,是研究基于数据的智能。其定义是:人工神经网络是一种计算系统,包括软件和硬件,它使用大量简单的相连人工神经元来模仿生物神经网络的能力。人工神经元是生物神经元的简单模拟,它从外界环境或其他人工神经元取得信息,并加以非常简单的运算,输出其结果到外界环境或其他人工神经元。在进行法律案例归类分析判决中,可以运用人工神经网络和遗传算法来解决问题。

2. 模糊比较推理

模糊推理以模糊集合论为基础描述工具,对以一般集合论为基础描述工具的数理逻辑进行扩展,从而建立了模糊推理理论。它是不确定推理的一种,在人工智能技术开发中有重大意义。

在 CBR 中,案例检索是推理的核心。郑永和等^[110]将模糊推理运用于案例检索的过程中:对于案例匹配失败的事实,模糊推理器根据已有的事实,采用计算最大隶属度的方法求出符合要求的知识框架,从案例库中取出尚未匹配成功的规则框架与已有事实框架进行模糊匹配,选出匹配度最大的事实,采用基于隶属度表示的不确定性推理方法和按匹配度排序的冲突消解策略,计算隶属度传播值,将成功匹配的规则和匹配次序号等标记及隶属度传播值等信息传给驱动解释部件、保存结论部件,并把事实组织成案例保存在案例库中。

3. 定性比较推理

定性比较推理是人工智能学科的一种推理方法,通过对系统的结构、行为、功能及它们之间的因果关系进行研究,来探索人类常识(定性)推理机制,从而有效地完成各项求解任务的一种跨领域的推理方法体系。定性推理可以有效地解决常识推理、反映动态特性的时态推理及反映直觉思维的因果推理等问题。定性推理理论产生以来,研究人员提出了多种建模和仿真实论,其中代表性的主要有以下几种:基于流概念、以组元为中心的 Envision 方法,以过程为中心的定性过程方法,以约束为中心的定性仿真方法 QSIM (Qualitative Simulation for Imprecise Model, QSIM) 及基于因果关系的因果推理方法^[111]。

4. 基于规则的推理

规则是指导人们在特定情况下作为或不作为的行为规范。基于规则的推理,

即把某一规则用于一个具体案件中,做出一个司法判决^[112]。典型的规则一般针对特定人的某一种行为,可以表述为:在 X 类条件下, Z 类人应该实施 Y 类或不实施 Y 类行为。也就是说,RBR是从一个已知的一般性的前提出发,结合具体事实,得出符合逻辑规定的结论来。可以把规则看成决策问题的大前提,事实看成决策问题的小前提,判决即为结论,这样就可以大致用演绎推理的模式予以概括。在实际司法审判中,推理则相对复杂,法官对于粗放的规则并不能全面把握,自由裁量将使本已宽泛的规则变得更加没有约束力。比如,面临一个案件时,法官往往需要选择规则,但是规则选择的过程实际上是一个难题,因为没有法律规定法官在具体的案件中必须如何选择适当的规则,也没有可靠的机制确保法官所选的规则就必然是适当的。虽然有明确的法律分类,如民法、刑法、行政法等,但这种分类在很大程度上只是为了国家立法和法学研究的方便。

5. 基于案例的推理

基于案例的推理符合人类思考问题的习惯,它是人类类比思维的一种推理方法的模拟,推理过程具有人类经验推理的特征。CBR的基本过程是:当遇到一个新的问题时,系统会根据决策问题的关键特征在原始案例库中进行检索,找出一个或多个与待求解问题最相近的案例,并在新问题解决时重用此候选案例的解决方法。当然,如果对选出的案例的解决方法不满意,还可以对它进行修正以适应待求解问题,最后把修改过的案例作为一个新的案例保存在案例库中,以便支持其他类似问题的解决。CBR以案例作为知识元,其知识获取和表示自然直接,并且本身具有自学习功能,这正是符合了人类类比思维的逻辑。

在CBR中,把当前所面临的问题或情况称为目标案例,而把记忆的问题或情况称为源案例^[113]。粗略地说,CBR就是由目标案例的提示而获得记忆中的

源案例，并由源案例来指导目标案例求解的一种策略。

CBR 有两种类型，即问题求解型和解释型。问题求解型侧重于对过去策略的匹配与修改，而解释型强调以旧案例对新案例做出评价与解释。无论哪一种，其推理过程均类似于人类经验类比推理，而且具有简化知识获取、通过直接获得提高求解质量与效率、适用于非计算推导的优点。因此，CBR 将是人工智能与专家系统设计的一种非常具有发展前景的方法。

6. 混合推理

法律推理本身是一种复杂的思维过程，并非机械地照搬法条，而是需要缜密的逻辑思维加上丰富的经验，才能保证推理结果合法、合理、合情。法律本体推理借助法律领域本体中定义的概念及概念之间的关系和依据法律规范建立的规则实现推理。以本体为基础的推理需要借助推理机实现，目前常用的推理机包括 Penet、Racer、FaCT++等。

在基于本体的法律案例推理方面，Adam Wyner 总结了两个著名的 CBR 系统——AS-CATO 和 IBP。

在 AS-CATO 中，案例使用因素和最终判决方相关联。其中，因素即案例的特征，用于决定最终判决方。案例可能有多个支持其的因素，每个因素均有 Side 属性表征其所支撑的一方（原告或被告）。例如，若因素 1 的 Side 为原告，即表示若原告持有因素 1 则对原告有利。这样，可以大致将因素分为原告因素和被告因素两部分。在基于案例比较的案例推理过程中，AS-CATO 将正在进行比较的两个案例的因素划分为 7 个集合，每个集合表示了其若非空对案例判决的影响（即判给原告和被告）。例如，第一个集合包含在两个案例中均存在的原告因素，如该集合非空并且其余偏向被告的集合为空，则倾向于判给原告；若

偏向原告和被告的集合同时非空，则需采用论据方案^[114]的方法进一步推理。

在 IBP^[115]中，使用了领域模型的逻辑结构对议题（即法律推理过程中的中间概念）进行分层组织，较低层议题的成立是与其直接关联的较高层议题成立的条件。案例因素并不是这一逻辑结构的直接组成部分，只用于支持底层议题，也具有 Side 属性。另外，因素对于最终结果影响的强弱也被考虑在内，据此将因素分为淘汰性因素和弱因素。在案例比较的过程中，没有使用 AS-CATO 中采用的因素划分的方法。

通常，在 RBR 中，规则的基本形式如下：

```
if<conditions>  
then<conclusion>
```

其中，<conditions>表示一条规则的条件，<conclusion>表示这条规则的结论。规则的条件之间通过逻辑连接词（如 AND、OR、NOT 等）相互关联，这就形成了一个逻辑函数。当充分条件得到满足时，就可得到结论，即该规则被激发。

很多专家系统都是基于 RBR 建立的，典型的专家系统包含有以下几个关键部分：规则基础、推理机、内存、解释机制。推理机使用知识（规则）解决问题。有两个主要的推理方法：反向推理和正向推理，反向推理是由目标所引导的，而后者则由数据引导。

Trevor Bench-Capon 等^[116]提到了一个基于规则的较早的法律推理系统，由 Anne Gardner 设计，主要领域为美国合同法，系统目标是“指出议题”，即对于给定的案例，识别出案例中容易的和困难的法律问题，对容易的问题给予解决。系统包含从合同法中获得的规则集及从上千个合同案例中提取的 385 个原

则。这些规则是连贯的，规则集又包含了补充的解释规则集，它们从判例法、常识和专家意见中提取。Gardner 认为，简单的问题是那些得到一个结果的问题；而困难的问题或议题，或是没有答案，或是由于匹配多个规则而得到多个结果。一些议题可以使用通过判例法和常识规则赋予规则优先级的启发式程序解决，余下的议题则需法官自行推理。

随着 RBR 和 CBR 模型及其相关的人工智能技术的日趋发展，加之对两种推理模型各自优缺点的认识，将这两种推理模型相结合成为了一种趋势。法律推理和医药领域成为了将规则和案例相整合方法的重点应用领域。Jim Prentzas^[117]总结了一些将 RBR 和 CBR 相整合的系统，其中的法律系统有 CABARET、DANIEL、GREBE、IKBALS III、SHYSTER-MYCIN。这些系统中的规则通常是基于法律法规，而案例是基于以往的可靠案例。在法律推理领域将两种推理相结合的趋势是源于“开放结构”规则条款的存在，规则可能有未被声明的先决条件及定义中的意外或循环。

C.R.Marling 等^[118]将 RBR 和 CBR 运用于日常营养餐规划的系统中，先由 CBR 系统和 RBR 系统分别独立执行相同的规划任务，其中 CBR 系统是通过查找匹配并借鉴先例中解决办法进行营养餐制定，而 RBR 是通过对一些预先设定的模式规则进行匹配来建立营养餐谱，然后结合分析方法评估 CBR 和 RBR 系统的餐谱制定结果，并实施整合。

Jim Prentzas 将 RBR 与 CBR 相结合的系统以图 2-4 所示的分类方式加以总结。这些系统有些是应用于法律推理领域，其他的一些则不是，在图中，只将应用于法律推理的系统在相应的类别后标出。其中，协作式处理是指构成的组件之间紧密地相互影响，在面向协作的系统中，部件相互协作（通常通过推理步骤的交叉），以得到其最终的结果；在面向协调的系统中，每个部件得到各自

的结果（这些结果通常互不相同），这就需要结果的协调，以得到最终的结果。

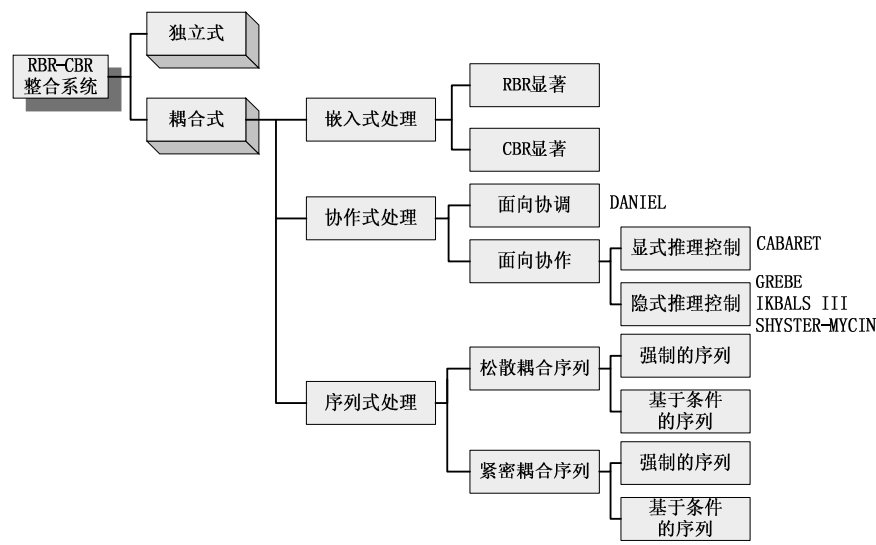


图 2-4 RBR 与 CBR 相结合的系统的分类

2.4 文献述评

通过 2.1 的分析可以看出，目前，对法律知识表示与抽取的研究主要集中在以下三个方面：第一，从本体的角度提出了法律本体的概念；第二，分析了法律本体的构建模式；第三，以文本挖掘为基础的信息抽取与标注算法。结合文献研究，对本书主要研究问题的价值及研究面临的问题述评如下。

2.4.1 研究问题的价值

通过综述可知，基于两种推理方法集成的量刑决策支持研究，在司法审判领域来讲，是一个全新的问题，具有一定的学术价值和应用价值。

1. 学术价值

从学术价值的角度讲,基于本体的刑事案例库的构建过程包含了三层学术意义。

一是将哲学中的本体概念用到了刑事审判这样一个新的领域,是一个方法与新领域的结合,是一个应用创新。具体来讲,把一个比较复杂的、混乱的领域知识,用一个方法进行了规则化、规范化,使其成为统领领域信息的总的知识框架。

二是将几种学科进行了结合,突破了原来决策信息的瓶颈,找到了一条基于文本的决策的信息转化途径。将半结构化的数据经过刑事案件本体的纲领化,进而获得概念及其相关互关系和属性值,并用恰当的方法将其从混乱的文档中抽取出来,将大量的“沉睡的”的数据,转换成有活力的信息;经过从描述性文字到规范性的转换,并利用数据仓库技术,将其存放在可以挖掘的数据库里面,形成可供决策的“精细”的决策基础信息,它可以直接与数据挖掘技术相衔接,进行决策支持。

三是管理决策理论的实证研究与现代软件技术的结合,促进了决策方法的研究,为相关领域的决策提供了方法。研究的全过程贯穿了应用现代成熟的软件进行实验。实验中可以知道,本体的构建和依赖本体对案例文档集进行抽取得到的案例库,实质是刑事案件的全集案例库,它包括了与案件相关的全部有用信息,可以用于刑事审判领域方面的分析。当然,在决策推理中只用到其中一部分信息,所以在后面的研究中,整个案例库是一个为了研究目的逐渐缩小、聚焦目标的过程,其集合更符合研究目的的要求。

2. 应用价值

从应用价值的角度讲，基于本体的刑事案例库的构建过程包含了三层应用意义。

第一，找到了比较好的对法律知识规范化的方法。用本体将刑事案件知识进行规则化表示，形成可共享、可借鉴的成果，供司法审判领域使用。

第二，为量刑决策找到了可靠的信息支撑。将基于多变量决策的决策问题要素信息化、规则化，使决策主体有条件更加便捷地得到这些信息，并对决策进行支撑，刑事案件本体、刑事案件抽取标注 OWL 模板、刑事案例库构成都对该领域有借鉴作用。

第三，为决策信息化建设提供了思路。从信息技术服务领域决策来讲，为从事信息技术但缺乏领域知识的技术人员，提供了解决数据到信息的技术方案，使服务审判一线科学决策成为可能。

2.4.2 研究面临的问题

1. 刑事案件数据表示

刑事案件表示作为法律知识表示的组成部分，受到领域知识的影响，更受限于适合本国法律特征的本体知识组织方式的影响。国外在法律知识表示的方面，从通用本体到领域本体研究了很多，甚至提出了基于知识的法律本体，在法律整体知识表达的理论层面相对成熟。相比较，国内这方面的研究显得不足。同时，国内外基于本体表示法律知识，还没有具体到针对特定司法领域中的一类案件的表示。对刑事案件信息的抽象化、规则化研究可以基于我国案件的文

档架构、信息特征及国外法律本体建立模型，进行刑事案件本体的建立，达到刑事案件知识规范表达的目的。

2. 刑事案件信息抽取

刑事案件信息以半结构化文档存在，以文本挖掘为基础的信息标注与抽取受到文本特征和关键词的影响，以前由于对文本特征缺乏规范化表示，自动标注不能实现全信息抽取，手工标注不能实现对大量数据信息的及时抽取。但前人的研究为本书研究提供了思路，在以刑事案件本体和刑事案件本体实例化的基础上，对半结构化的刑事案件文档集，按照一定的抽取规则，对文档中的概念、实例和关系进行抽取，使得量刑决策赖以的半结构化数据得以规则化。

3. 刑事案件案例库

刑事案件数据库存储的大都是结构化的案件信息，这些信息大都是审理中的程序信息，刑事案件的大部分实体内容都蕴含在刑事判决书中。刑事案件表示、抽取的目的就是将这些实体信息和原来的结构化信息共同组成反映刑事案件全貌的规则化信息。我国近几年有的《人民法院案例选》可以进行查看，美国自 1790 年以来的完备的电子版判例集也仅可以检索，都不可能支持以它们为历史案例的数据挖掘。建立可供挖掘的刑事案例库，无论是从信息技术角度，还是司法审判领域，都是一个新的内容；前者从可行性和路径上做了大量研究，后者从决策需求角度提出了强烈的愿望。根据刑事案件特征和决策需要，进行数据结构的逻辑设计，以及数据存储的物理设计，实现支撑决策的信息瓶颈的突破，是决策推理的前提。

4. 相似刑事案件

在基于可挖掘数据库作为决策基础信息之前,大多数据分析相似案例都是从文档集的相似度入手,有语词、句子、概念等的相似度来代表文本的相似度,进而代表文档内容的相似度。无疑这种研究,处在数据挖掘信息瓶颈时是可以接受的;但现在基于半结构化文本决策信息的瓶颈已经不在,对此类研究继承的价值就不大了。基于以结构化数据存储为特征的案例库或数据仓库,分析其中表达的案件的相似程度,用于推理决策模型训练和使用之中,并将相似案例的选择、修正和入库动态化,保证决策模型对决策对象的实时适应性。关键问题是针对结构化案例的数据量大、属性多的特点,必须将案例库进行分类和细化,针对决策要求,按照不同的刑罚类型,将其快速地检索,这就需要将大量的属性进行优化,以提高检索速度,实现推理决策需要的相似案例库,达到缩小案例空间的目的。

5. 推理方法集成

从 2.2 中可知,国内外对法律知识的推理研究,都是将实现人工智能的计算机算法用于法律领域,研究全面、内容丰富。有基于规则应用的规则推理、基于案例复用的案例推理和二者结合的混合推理。国外法律领域已经有将规则推理与案例推理进行混合使用的理论框架,其中的协作式处理是在推理过程中,将二者交叉使用,但要得到科学的决策,必须与领域决策信息特征相结合。另外,国外司法审判领域的推理重点放在对证据采信与支持审判的哪一方上面,而本书研究的推理重点是放在利用历史案例支持目标案件的判决上。推理模型在应用前,能根据刑事相似案例库进行有效的训练,特别是将决策树、神经网络算法与领域知识结合,使规则推理算法、案例推理算法实例化,将两种推理在同一决策中恰当使用,以发挥各自优势,达到最终目标。

鉴于以上综述中刑事案件量刑决策研究的不足及决策推理的理论层次不够,本书将本体理论、数据挖掘方法、决策理论等运用到刑事案件量刑决策的研究中,对刑事案件知识创新表达方式、刑事案件决策信息抽取、相似刑事案件选择、规则与案例两种推理方式集成,以及辅助量刑决策的系统原型进行深入的理论解释和探讨,通过实证方式提高其研究层次和理论高度,使其决策方法更具有科学性和普适性。

基于本体的刑事案例库构建研究

法官量刑决策的基础数据是多年来积累的判决书，它蕴含着案件的案情、法律依据、判决结果和法官的审理经验，由于其半结构化的特点使得不能同时使大批量的判决书文档集直接用于支持法官决策，必须将其结构化、规则化。本章研究的问题是半结构化判决书结构化、规则化，解决法官量刑依赖的决策基础信息问题。研究思路是以本体为核心，从刑事案件本体模型、实验和评价的视角对结构化方法进行论证，通过实证证明该方法的有效性。最终实现将半结构化判决书文档集转化成反应案件全部信息的案例库，满足案例指导制度背景下人民法院案例系统的建立^[119]。

3.1 构建刑事案例库的条件分析

3.1.1 刑事案件文书特点

刑事案件文书包括起诉书、判决书等法律文书。它是法律领域内的一种专用文书,除了具有与其他文书共同的特点外,还有其自身的合法性、固定性和强制性等特点。特别是固定性特点,使得刑事案件判决书得以全部实体信息抽取,区别于其他研究中只能对文本的摘要信息进行提取。

1. 刑事案件文书的合法性

刑事案件文书的合法性主要表现在三个方面。首先,必须依照程序制作。因为法律文书在诉讼活动中使用,它的出具使用直接反映着诉讼活动的进展,所以必须依照程序法的有关规定制作。如刑事案件的处理,从案件的立案、侦查、破案、报捕、批捕、审讯、移送起诉到审查起诉,其中,诉讼的每个环节都需要制作相应的文书来作为进行某项诉讼案件活动的文字凭证,都必须按照《刑事诉讼法》有关条款规定制作相应的法律文书。在公诉案件刑事诉讼活动中,只有按上述规定的特定阶段予以制作才属合法。

其次,文书时限的合法性。时限的合法性也是法律文书制作的合法性的一个特定要求。如我国刑事诉讼法中对采取强制措施的取保候审、监视居住也规定了最长期限,即对犯罪嫌疑人、被告人取保候审最长不得超过12个月,监视居住最长不得超过6个月,逾期如不对上述强制措施法律文书予以撤销或变更即属违法。因此,司法机关在处理案件的某项活动中必须按照诉讼法的规定遵守特定的时间,当事人行使某项权利也必须遵守特定的时限。

最后,法律手续的合法性。法律文书的合法性还表现在某些文书的使用必

须履行特定的法律手续。如刑事案件对犯罪嫌疑人扣押物品的，扣押物品清单中在写明扣押的物品名称、数量之后，须由被扣押的人签字或按手印认可。如不履行此项手续，不仅不能发生法律效力，并且该项活动也属违法行为。

2. 刑事案件文书的固定性

刑事案件文书的制作有固定的格式，这是在长期的实践中不断创造、改进而形成的。它既能保证法律文书的完整性和严肃性，容易全面有力地发挥法律文书的作用，又能保证制作时简易方便，具有科学性。其具体表现在以下几个方面：

首先，文书结构的固定化。不同种类的法律文书，其行文表述的结构大都有固定的格式。一般来讲，都具备首部、正文、尾部三部分内容。首部大都由文书标题、文书编号、当事人的身份事项、案由、案件来源等项目内容组成，并按上述次序排列。正文包括案件事实、处理理由、处理决定（意见）三项内容，是法律文书的核心内容。尾部一般由交代有关事项、签署、日期、用印、附注事项等内容组成。

其次，文书用语的规范化。各类法律文书的各部分内容的表达，都有规范的固定用语，书写该项目时只能如此表述，没有丝毫的变通余地。如人民检察院起诉书在写明案由及案件来源时，须使用如下固定用语表述：“被告人×××因××一案，由××侦查终结，于×年×月×日移送我院，经依法审查表明。”又如，人民法院的一审刑事判决书尾部向被告交代上诉权事项时，也必须用如下固定文字表述：“如不服本判决，可在接到判决书的第二日起十日内，通过本院或者直接向×××人民法院提出上诉，书面上诉的，应当提交上诉状正本一份，副本两份。”

最后,文书事项的要素化。各类法律文书在某些特定项目内容的表述中还须符合其要素规定,不可残缺不全。如在表述当事人身份事项时,应写明姓名、性别、出生年月日、民族、籍贯、工作单位、职业、住址等要素。刑事案件被告人还应写明何原因、何时间曾被拘留、逮捕、现羁押何处等要素。为了加强文书撰写的规范,我国相继出台了规范法律文书制作的相关标准文书格式,如《公安机关刑事法律文书格式》、《刑事检察文书样式》、《直接受理刑事案件文书样式》、《人民检察院刑事诉讼法律文书样式》和《法院诉讼文书样式》等。这些新格式样本的颁布,标志着我国法律文书学发展的进一步完善,它对于提高法律文书制作质量及提高办案质量都是一个有力的保证;同时,使本研究中,以判决书为研究对象,用本体的方法将其结构化、规则化成为可能。

3. 刑事案件文书的强制性

法律文书是法律实施的重要手段,因而它的实施必须依靠国家的强制力来保证其执行的有效性。对于一些具有执行意义的文书,这种法定的强制力就表现得更为明显。如本书研究的刑事判决书就是法律强制性较强的法律文书,二审刑事案件人民法院对被告人一经宣判,判决书就发生了特定效力,判处徒刑的,需立即收监服刑;判处死刑的,经过死刑复核程序,必须押赴刑场,执行枪决。

3.1.2 刑事案例现状分析

国家权威案例载体不能满足实际审判决策的需要。截至2013年,我国权威的案例载体主要有《国家最高人民法院公报》、《中国审判要览》、《人民法院案例选》和案例版的《人民司法》,这些案例涵盖了“问题提示”、“裁判要点”、“案情”、“审判”、“评析”五个部分。这些案例只是特定案件的分析,数量很少,

缺乏对所有案件类型的指导，没有系统性，而且是在少数人的判断中生成的。因为是纸质的，从检索和支持决策的便捷性来说就无从谈起了。

案例数据的海量化与信息“沉睡”。随着信息技术的飞速发展，刑事案件判决书也以电子文档的形式在数量上不断增长，仅××法院已经有140多万件刑事案件判决书。在信息系统的支持下，使法官可以方便地检索相关案例，但对案例本身蕴含的司法理念和审理经验的分析与挖掘，都没有提供一个可操作的环境。让人们感觉到的是数据很多，信息却很少，大部分有用的、有价值的信息处在“沉睡”之中。

案例信息的混乱与知识组织缺乏。尽管判决书有一定的规则，从标题到最后的落款都有一定的规范，但其最多是一种半结构化的文本，它具有半结构化数据的特点：（1）包含有隐藏的模式信息。虽然具有一定的结构，但因为数据与结构混杂在一起，所以并没有明显的模式定义。（2）结构不规则。半结构化的数据集合可以由不同结构的元素组成，或者某些情况下用不同类型的数据表示相同的信息。（3）类型约束不严格。数据在结构上没有遵循相同的规则，导致对数据没有严格的约束。对此，在我国司法审判历史上，还没有一套完整的组织和管理案件信息的框架。

综上，我国案例系统的建立，需要从案件信息的表达开始，司法审判知识的组织是支撑量刑决策的前提。尽管当前的案例文档不能满足决策需求，但至少为量刑决策研究提供了半结构化的研究对象和研究的起点。

3.1.3 刑事案件本体构建难度

构建刑事案件本体，规范司法审判领域知识表示，无论在统一司法审判领

域的知识表达,还是从中探索司法审判规律支持量刑决策,都具有很强的必要性。但要真正做好这一研究工作,还存在着一定的难度。

首先,刑事案件本体建模可参考的文献不多。本体建模有不少的文献,但司法审判领域的本体建模,特别是国内在这一领域的建模还基本没有。同时,我国完整的法律体系建立较晚,这一领域的专家更多专注的是实务研究,从事案例库系统建立的人员往往是信息技术人员,缺乏支持实际需求的框架设计和知识组织经验。

其次,刑事案件本体应用的难度在于交叉学科的共同作用。刑事案件本体的应用无疑有很好的前途,但使其真正发挥应用的作用,需要领域专家、管理科学专家及信息技术的共同完成,在决策学术研究和应用研究都有一定的难度。一是需要有丰富的基础数据;二是研究人员要具备司法审判领域知识,同时需要决策理论知识;三是需要有数据挖掘和实验软件的应用能力。这个过程是一个漫长的系统,需要付出努力才能实现。只有做好应用领域与信息技术的深度融合,才能达到预期目的。

最后,刑事案件本体的学术价值在于静态和动态信息的结合。刑事案件本体建模将法律知识的概念、概念属性和概念属性值以罪名为类进行了有机组织,它反映的是这类案件的静态知识,是这一罪名共有的知识结构;将刑事案件本体用于刑事案件信息的抽取和案件推理,这是刑事案件的动态信息,是动态知识。其学术价值在于,在刑事审判这一领域,使用本体的静态知识进行动态推理,是智能决策的基础。

因为上述的难点,本研究以本体知识表达应用、半结构化数据信息提取和支撑决策案例库构建组成本章内容,以探求在这一领域的创新应用。

3.2 刑事案件本体化框架设计

首先，从知识表达理论出发，将知识本体表示方法用于刑事审判领域，使其知识结构化；其次，将刑事案件本体用于半结构化刑事案件信息抽取，实现决策案例数据到决策案例信息的转化；最后，将抽取的刑事案件信息，用数据仓库的形式表示成可供挖掘的案例库。具体刑事案件本体化研究框架如图 3-1 所示。

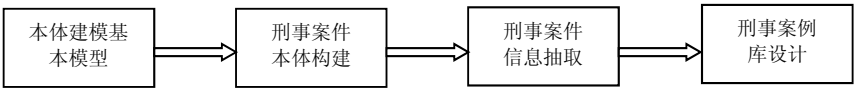


图 3-1 刑事案例库建立模式框架

本章研究问题是将本体的方法用于刑事审判领域的知识规范化表达，利用刑事案件本体为统领抽取出刑事案件文档集中的有价值信息，将有价值的信息进行转换与存储，形成可挖掘的刑事案例库，使之成为支持决策的基础数据库。其核心是本体方法在司法审判领域的应用和利用本体方法获取并存储决策信息的研究。

3.2.1 本体建模基本模型

为研究如何使用本体组织知识，Perez A G 等^[120]用分类法组织 Ontology，归纳出 5 个基本的建模元语，即概念（表示的是对象的集合，其定义一般采用框架）、结构、关系（在领域中概念之间的交互作用，形式上定义为 n 维笛卡儿积的子集： $R:C_1 \times C_2 \times \mathbf{L} \times C_n$ ）、函数（形式化的定义为 $F:C_1 \times C_2 \times \mathbf{L} \times C_{n-1} \rightarrow C_n$ ）、公理（永真断言）和实例（从语义上讲实例表示的就是对象）。本体正是通过这些建模元语，来组织现实世界的知识。

描述逻辑以其在语义、可判定性及面向对象的分类表示等方面的优良特性成为表达本体的有力工具,因此运用描述逻辑建立本体语义模型是本体形式化的基础工作。具体讲,它作为一种用来描述概念和概念层次关系的知识表示语言,尽管没有一阶逻辑的表达能力强,但推理复杂度是可预知的,适用于推理功能的应用;同时,它所依赖的语法容易转换成 XML/RDF 形式,适于计算机在网络环境中逻辑判断与知识共享。

本体模型通常是一个 4 元组,根据实际需要可以增删元组,记作: $O = \langle S, CS, RA, IS \rangle$ 。其中, O 为本体; S 是定义概念的构造算子集; CS 是概念集,包括原子类概念 AC (简称原子类)、原子属性 AP 及复合概念 CC , 记为 $CS = \langle AC, AP, CC \rangle$, CC 中的概念由 AC 中的概念通过 S 中的构造算子获得; RA 包括概念关系集及公理; IS 为关系断言及概念断言集,用来表明实例。

关于本体工程的建模方法,当前,在知识工程界比较成型^[121]的主要有: Uscholddede & King 的“骨架”法、Mike Gruninger 和 Fox 的“评价法”、Bcrncras ctal 方法、Methontology 方法、SENSUS 的描述方法、Mario Bunge 的领域本体建模方法论。

领域本体构造过程如图 3-2 所示。它是一个螺旋状的结构。起点是选择通用核心本体,任何大型的通用本体(如 Cyc 本体)、词汇—语义网(如 WordNet、GermaNet)或者领域相关的本体(如 TOVE)都可以作为这个过程的开始;选定通用本体后,法官必须确定用于抽取领域相关实体的文本,提炼出该领域的重要概念和主要的关系,编制一份“术语集”,并对每个术语注解相应的自然语言描述;为了让机器能理解现实世界中的概念和关系,必须对本体进行形式化的编码,一般采用较多的是基于描述逻辑的形式化编码方法,并用本体描述语言(如 OWL)来编写本体,为本体共享和重用提供规范的文档依据;对本体

的评价指标主要有：本体正确性、一致性、可扩展性、有效性和本体的规模及描述能力等。

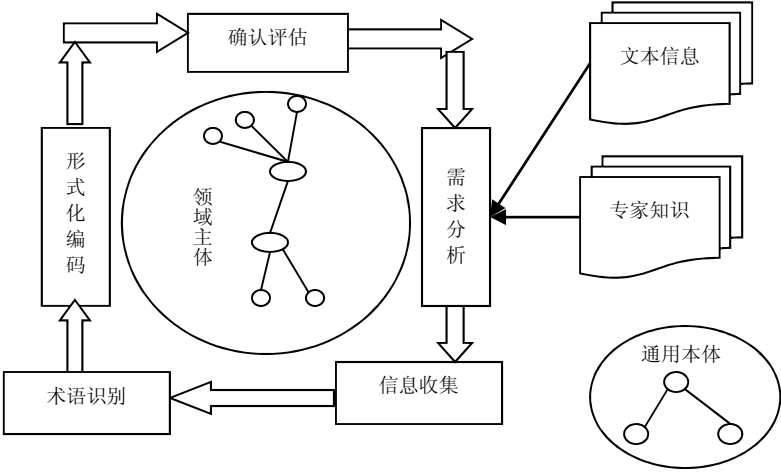


图 3-2 领域本体构造的原型模型

3.2.2 刑事案件本体构建

1. 刑事案件本体模型

本研究在其描述逻辑方式基础上，结合刑事案件相关知识的概念清晰、层次关系强的特点，提出了五元组的表示法的刑事案件本体模型，即 $O = \langle T, X, XD, AA, TR \rangle$ 表示刑事案件本体。其中， T 是术语集，包含类术语集 TC 、数据属性术语集 TD 和对象属性术语集 TO ； X 为实例集； XD 为实例声明集，用来声明术语的实例； AA 为属性分配集，包含用于确定对象属性所涉及类的对象数据分配集 AAO 和用于给每个类术语确定其所含有的数据属性和对象属性的类术语属性分配集 AAC ； TR 为术语约束集合。

2. 刑事案件本体设计

我国的刑事案件罪名有 400 多个, 其中案情和量刑相差很多, 如果从刑事案件的角度建立本体模型, 其中的共性概念及概念间关系比较宏观, 不能将真正需要的案件中的概念和信息组织进去, 为此本书提到的刑事案件本体是以罪名为类的, 如故意伤害罪类本体。

按照刑事案件本体模型 $O = \langle T, X, XD, AA, TR \rangle$ 和描述逻辑要求, 要建立一个刑事案件本体, 将涉及刑事审判相关的要素, 按照良好的概念层次和知识结构组织起来, 为司法审判领域探索规范审判知识表示途径。遵循图 3-2, 整个构建过程由以下 5 个环节组成。

(1) 确定刑事案件通用本体。以刑事审判领域相关本体或概念集, 作为通过本体, 作为设计的起点。

通过文献检索, 在我国司法审判领域尚没有通用本体, 《新编常用法律词典》中的“刑法部分”, 对刑事审判中的 86 个主要概念及其关系做了比较好的界定, 本书将其作为构造刑事案件的通用本体, 同时结合领域知识和《人民法院信息化建设规范》中的词条, 一并作为本研究的起点。

(2) 对刑事审判领域相关实体的文本进行整理、确定, 形成刑事案件基础“术语集”。

刑事案件基础“术语集”。以刑事案件通用本体为基础, 通过对人民法院的刑事案件判决书进行语词分析, 并分析刑事案件判决书的行文特征和叙述内容、结构, 并进行语词分析和聚类, 将构成刑事案件的实体内容分为 10 类, 包括: 审判法院、判决书编号和公诉机关; 被告人及其辩护人、法定代理人(仅限于无诉讼行为能力的被告人) 信息; 起诉书信息及其中的指控内容、案情等; 被

告人辩护意见；经审理查明案情；公诉方宣读和出示的证据；法院予以确认的证据；对被告人辩护意见的采纳情况；依据法条及判决结果；合议庭组成及判决日期。这 10 类信息作为刑事案件本体设计的基础“术语集”。

(3) 从基础“术语集”中，提炼出刑事审判领域的重要概念和主要关系，编制成“术语集”，也就是本体。

在此基础“术语集”上，通过聚合的方法，进行刑事案件本体的概念类、子类设计和属性类设计。其中，属性类包括对象属性和数据属性。子类设计是为了细化概念，同时要验证本体类之间的逻辑关系（父-子类关系）是否是前后一致的：例如，类 *A* 是类 *B* 的父类，同时类 *B* 和类 *C* 是等价类，那么若设计类 *C* 是类 *A* 的子类就不符合概念的一致性，在进行本体实例化之前需对此作出检查。

本体概念类及其关系的设计。本体概念类的设计是刑事案件本体设计的核心，通过概念类的定义和之间的语义关系确定构成刑事案件实体信息组织框架。结合领域专家意见，本书使用法律机构、起诉书、合议庭和判决书 4 个类建立本体的顶层概念框架。其中，法律机构包含法院和检察院两个子类；判决书类用于组织法院判决的相关信息；起诉书类中存放检察机关向法院起诉时所提出的指控信息；合议庭是法院审判案件的基本审理组织，该类中包含负责审理案件的审判长、审判员、人民陪审员、书记员等相关信息。上述四个顶层类之间的关系如图 3-3 所示。四个顶层概念类的组织以判决书类为核心，图中的语义关系 A 表示判决需参考起诉书所包含的指控信息；语义关系 B 表示起诉书是由检察院撰写并提出指控的；语义关系 C 表示判决书是由合议庭对案件进行审理并做出判决；而案件的审判机构是法院，即语义关系 D；箭头 E 表示合议庭由法院组织。

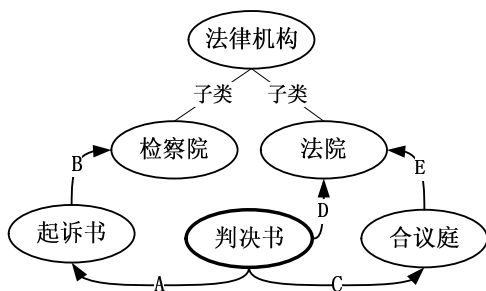


图 3-3 刑事案件本体顶层概念框架

在完成对刑事案件本体的顶层概念建立之后，使用法律角色、案情、证据、审判结果和依据法条 5 个类进一步细化本体的组织结构并丰富其中所包含的语义信息。

法律角色类的组织结构和语义关系如图 3-4 所示。法律角色类包含被告人、辩护人、法定代理人这 3 个子类，其中，被告人类又有子类起诉信息_被告人。起诉信息_被告人子类用于存储在起诉书中抽取的被告人信息，包括其所被指控的罪名；被告人、辩护人、法定代理人子类与判决书中的相关信息相对应。语义关系方面，图中的 a 箭头表示判决书中包含有被告人信息，在有些情况下，一份刑事判决书会含有多个被告人；b 箭头表示被告人可以委托辩护人为其辩护，一般情况下，有一位或两位辩护人，也有被告人自行辩护的情况；c 箭头表示无诉讼行为能力的被告人，如 18 周岁以下被告人，由其法定代理人进行诉讼；另外，对于判决书的每位被告人，均会得到判决结果，如 d 箭头所示。

案情类的组织结构和语义关系如图 3-5 所示。案情类包括起因、经过、结果、案情相关类这 4 个子类，前 3 个子类用于存储刑事案件的起因、经过、结果等相关信息；案情相关类主要用于对经过子类和结果子类进行描述。由于这些用于表述的子类与经过子类和结果子类之间并没有父-子类的关系，所以设计案情相关类对它们进行组织，包含袭击部位、作案工具、伤情等级、从轻情节、

从重情节这 5 个子类。语义关系方面，图中的 e 箭头表示公诉机关在起诉书中所指控的案情；f 箭头表示合议庭在对案件的审理过程中所查明的案情，上述两个语义关系对应案情类的不同实例。g、h 箭头表示案件经过中的袭击部位和所使用的作案工具；i、j、k 箭头表示故意伤害案件所造成的伤情等级（包括轻伤、重伤、死亡等）、从轻情节（如救治、自首、赔偿等）和从重情节（如逃逸等）。其他类组织结构与语义关系略。

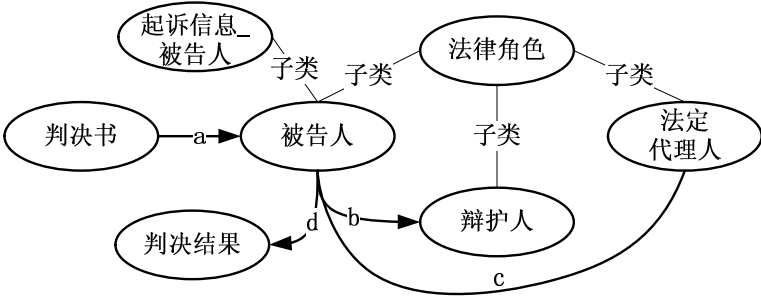


图 3-4 法律角色类

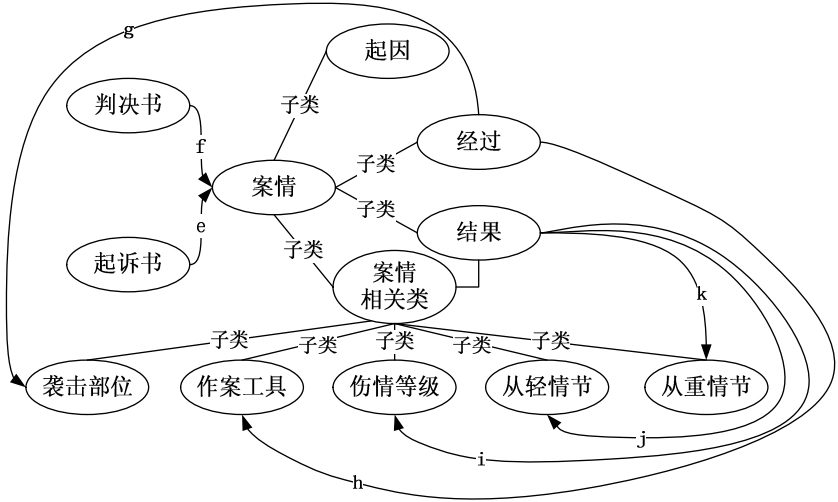


图 3-5 案情类

本体属性的设计。本体属性用于描述本体，分为对象属性和数据属性。对象属性表征类之间的语义关系，数据属性用于描述类的特征。刑事案件本体的对象属性，即四个概念和五个类之间的语义关系，在前文中进行了设计，这里重点介绍数据属性的设计。通过对刑事案件判决书文本的理解和抽象，使用数据属性描述类所包含的信息。如下一段文字为某刑事判决书中的被告人信息部分，以此为例，介绍数据属性的设计，以及使用数据属性组织信息的方式，使非结构化的刑事判决书趋于结构化和规则化。

被告人吕××，男，33岁，1973年3月16日出生于×市，汉族，大学文化，×市公安局××分局××派出所民警，住本市××区××楼××号（户籍所在地：本市××区××号）。因涉嫌犯刑讯逼供罪于2004年11月22日被羁押，同年12月7日被逮捕。现押在×市看守所。

可以用下面数据属性组织方式，来描述“被告人”这个类的特征。

被告人姓名：吕××

被告人性别：男

被告人年龄：33

被告人出生年月：1973年3月16日

被告人出生地：×市

被告人民族：汉族

被告人文化水平：大学

被告人职业：×市公安局××分局××派出所民警

被告人住址：本市××区××号楼××号

被告人户籍地：本市××区××号

涉嫌罪名：刑讯逼供罪

被羁押日期：2004 年 11 月 22 日

被逮捕日期：2004 年 12 月 7 日

现羁押地：×市看守所

是否累犯：否

同理，可以对刑事判决书中其他类的特征进行细化表述成数据属性，如表 3-1 所示为故意伤害罪的数据属性设计，对刑事案件本体的类的数据属性进行了总结，从中可以看出数据属性是对类的特征的进一步描述。

表 3-1 数据属性设计

序 号	类 名	数据属性
1	判决书	审理状态；判决书编号
2	起诉书	起诉书编号；提起公诉日期
3	合议庭	审判长；审判员；代理审判员；人民陪审员； 书记员；检察员
4	法律角色	被告人子类中：被告人姓名；被告人性别；被告人年龄；被告人出生年月； 被告人民族；被告人住址；被告人户籍地；被告人文化水平；被告人职业；涉 嫌罪名；被羁押日期；被逮捕日期；现羁押地；是否累犯 起诉信息_被告人子类中：起诉信息_被告人姓名；起诉信息_指控罪名 辩护人子类中：辩护人姓名；辩护人单位 法定代理人子类中：法定代理人姓名；法定代理人年龄；法定代理人出生日 期；法定代理人民族；法定代理人职业；法定代理人住址；与被告人关系
5	案情	案情_被告人姓名；犯案时间；犯案地点；被害人 起因子类中：事件；先施暴方 经过子类中：施暴方式；受害方；受害人个数；攻击次数

续表

序 号	类 名	数据属性
6	证据	物证子类中：获取物证时间；归属；物品 人证子类中：证人证明时间；证人涉案关系；证人证明地点；证人证明内容
7	判决结果	罪名；刑期；其他附加
8	法条	法条编号；法条内容

需要说明的是，每个类的数据属性并非在所有的刑事判决书中都会出现，如有些合议庭包含了审判员，而另一些则为代理审判员。因此，依据不同判决书的实际情况，有些数据属性可以为空值。

综上，以故意伤害罪为例，将刑事案件本体的类、对象属性和数据属性的逻辑关系用图 3-6 表示，即刑事案件本体的总体框架。刑事案件本体类与属性的详细设计见附录 A。

（4）对本体进行形式化的编码，并用本体描述语（OWL）来编写本体。这样，才能使计算机理解现实世界中的概念及其关系，为本体共享和重用提供规范的文档依据。

仅仅通过表或图的方式表达刑事案件本体的类框架与语义关系是不够的，必须将其进行形式化编码，使得计算机能够识别刑事审判过程中的概念及其相互关系。本书用常用的本体描述语言（OWL）对上述本体进行了编写，其过程如图 3-7 所示。

在完成本体实例化过程之后，将各罪名的刑事案件本体存储为 OWL/RDF 格式的文件，便于后期的信息抽取使用和机器学习。这里 OWL 语言是通过 XML 组成的，原因是利用 XML 的独特特点：① XML 语法具有严格的语法规则。② XML 对于各元素的顺序有着严格的规定。③ XML 中的元素可以有多个属性值。XML 成为数据挖掘中数据源层的数据集成，即构建数据仓库。

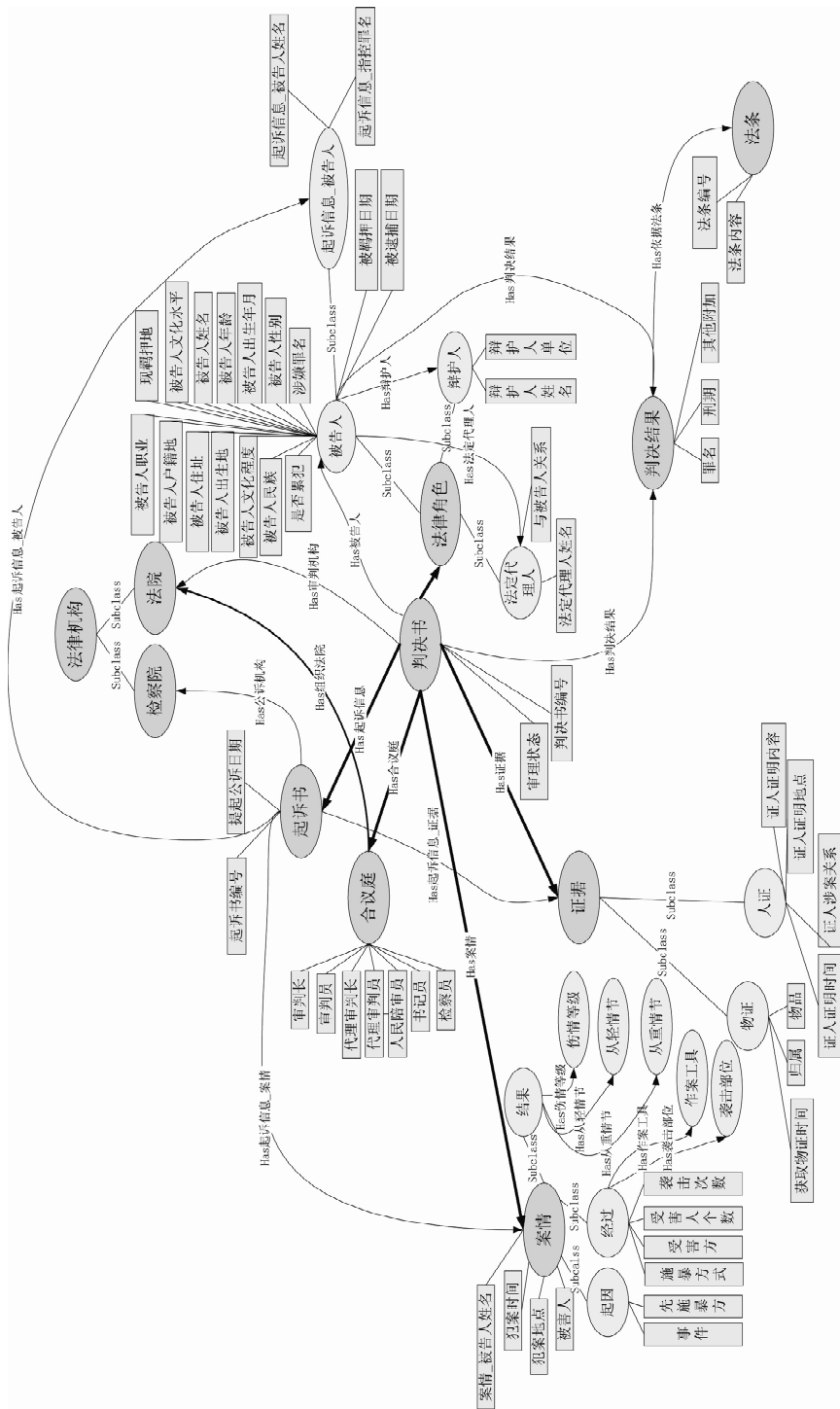


图 3-6 刑事案件本体的类框架与语义关系

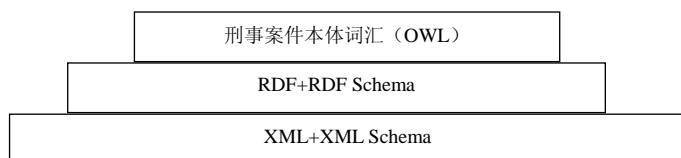


图 3-7 刑事案件本体形式化编码方法

限于篇幅，这里只列出对象属性中的 **Has** 公诉机关的 **OWL** 描述和法律机构类及其子类法院和检察院的 **OWL** 表述。以故意伤害罪为例，详细的刑事案件本体 **OWL** 文件见附录 B。

```

<!-- http://www.semanticweb.org/ontologies/2012/11/Legal
Ontology1.owl#Has 公诉机关 -->
<ObjectProperty rdf:about="&LegalOntology1;Has 公诉机关">
  <rdfs:range rdf:resource="&LegalOntology1;检察院"/>
  <rdfs:domain rdf:resource="&LegalOntology1;起诉书"/>
  <rdfs:subPropertyOf rdf:resource="&LegalOntology1;起诉书类中的
对象属性"/>
</ObjectProperty>
.....
<!-- http://www.semanticweb.org/ontologies/2012/11/Legal
Ontology1.owl#法律机构 -->
<Class rdf:about="&LegalOntology1;法律机构"/>
<!-- http://www.semanticweb.org/ontologies/2012/11/Legal
Ontology1.owl#法院 -->
<Class rdf:about="&LegalOntology1;法院">
  <rdfs:subClassOf rdf:resource="&LegalOntology1;法律机构"/>
</Class>
<!-- http://www.semanticweb.org/ontologies/2012/11/Legal
Ontology1.owl#检察院 -->
<Class rdf:about="&LegalOntology1;检察院">
  <rdfs:subClassOf rdf:resource="&LegalOntology1;法律机构"/>
</Class>
  
```

3. 刑事案件本体一致性检验

刑事案件本体一致性的评价，主要是对本体中的概念及概念之间的关系进行评判，避免概念关系相斥的情况出现。本书使用本体编辑工具 Protégé 4.1 中基于描述逻辑的推理机插件 FaCT++或 HermiT 进行概念一致性检验，推理完成后，在 Class 选项卡的 Class (inferred) 窗口中出现的类架构与所涉及的类架构相同，即通过测试。如图 3-8 所示，左图为使用 HermiT 推理机的过程，右图推理得到的满足一致性检验的类架构与所涉及的相同。

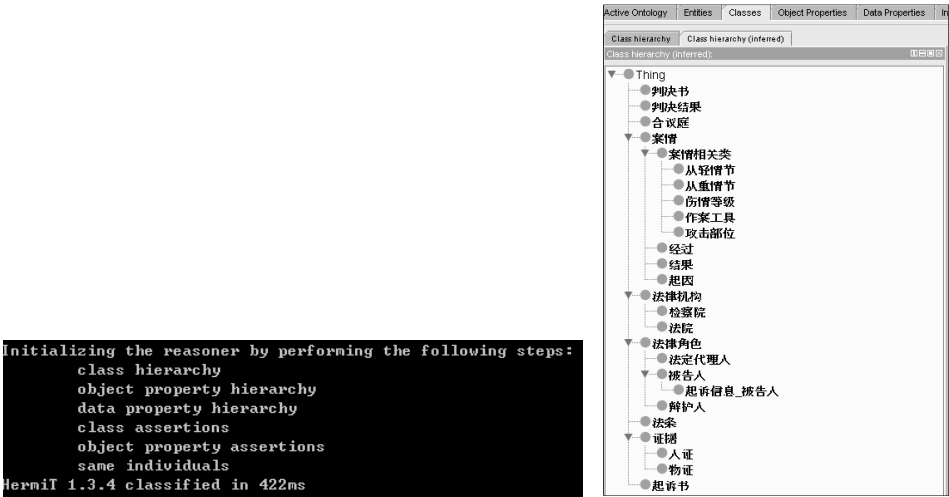


图 3-8 使用 HermiT 推理机验证一致性

对本体描述的一致性进行评判后，按照罪名对案件本体进行存储，形成以罪名为分类的刑事案件本体库，以此作为刑事案件信息抽取的输入条件。

3.2.3 刑事案件信息抽取

信息抽取技术是数据挖掘领域对半结构化、非结构化的文本数据进行预处理的关键技术，抽取的过程是指从文本中抽取指定的事件、事实等信息，并形

成结构化存储的过程。Cardie C 描述的信息抽取通用过程包括符号化和标注、句法分析、抽取、合并与模板生成五个部分。语义标注与信息抽取是密不可分的两个关联过程，是在一个领域本体的指导下为文档添加规范化知识表示的过程，包括类型标注和关系抽取两个部分。

目前已有的语义标注方法主要有三种，第一种是基于机器学习和关系元数据的传统方法，主要是对标注好的训练集进行训练，但在实际应用中需要定义匹配规则等来实现知识的转化；第二种是基于本体的方法，如 SemTag 利用本体中已有的实例来简化抽取过程中对概念实例的识别，这种方法较难捕捉实体间的关系；第三种是自然语言处理的方法，借助语法结构和句法分析实现实体关系的抽取，如 Artquakt、RelExt 等。

1. 刑事案件信息抽取模型

3.2.2 中设计并实现了刑事案件本体，它是刑事审判领域概念及其关系的知识本体组织，是对刑事案件中的知识抽象与表达。由于本体表示的规范性，为基于本体的刑事案件信息抽取模型建立打下了基础。领域本体与信息抽取的结合，是现在信息抽取的热点，大都把领域本体和信息抽取技术结合在一起，根据本体描述的概念、关系、层次结构和概念与关系之间的约束等生成抽取规则，然后再根据规则对输入的文档进行信息抽取^[122]。

刑事案件信息抽取标注系统总体框架如图 3-9 所示，该框架分为系统输入层、本体标注层及系统输出层三个层次。以特定罪名的案例文档集（刑事判决书）和已建好的对应罪名的刑事案件本体为系统输入层；在本体标注层，通过运用自然语言处理中的语法分析，结合模式识别方法，运用正则表达式，识别命名实体，如案例中角色、时间等信息，抽取识别出的实体进行类型标注；结合正则表达式构造规则抽取器，结合实体实例，进行实例属性值的抽取，并在

实例抽取后进行实例间关系的抽取。最终将案例组织为 XML 的格式，将案例文档集表示为标记的 XML 文档集。

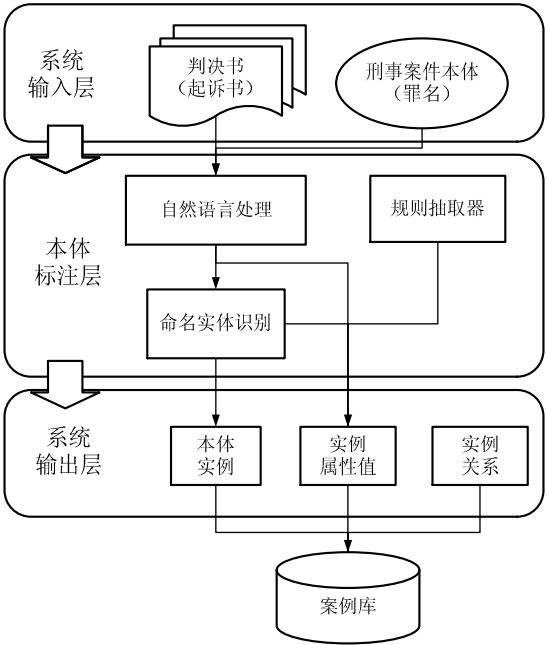


图 3-9 刑事案件信息抽取标注系统总体框架

2. 刑事案件信息抽取算法

刑事案件信息抽取算法，主要针对刑事判决书基本的写作规范和写作格式，结合自然语言处理的通用流程进行抽取算法的设计，主要分为篇章分析、句级抽取、词级抽取和标注四个部分，如图 3-10 所示。

第 1 步，篇章分析。依据刑事判决书的组成及内容分布，结合正则表达式将刑事判决书拆分为判决书基本信息、法律角色等六个部分，并标注。

第 2 步，句级抽取。对包含多项相似模式内容的部分，识别起始模式对其进行内容子块的划分。

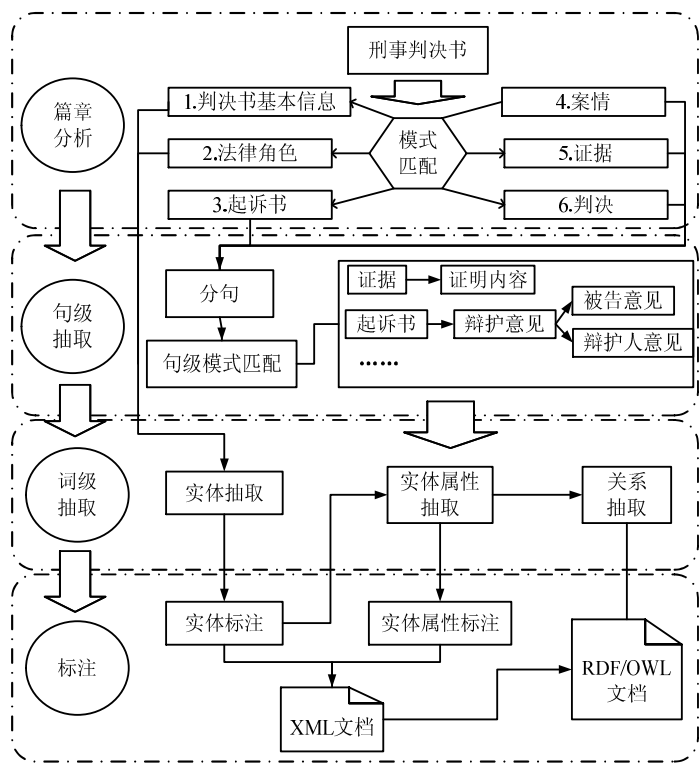


图 3-10 抽取算法

第 3 步，词级抽取。参照 3.2.3 节内容实现对命名实体、实体属性及关系等详细信息的抽取。

第 4 步，对抽取结果进行标注，实现本体的实例化。

表 3-2 中以“判决书”中“法律角色”部分为例，给出其信息抽取的算法，其他部分的算法略。

表 3-2 算法示例

For 文档集中的每个“刑事判决书”文档{
if 匹配“法律角色”的模式 {
抽取“法律角色”子类
if 匹配“被告人”的模式 {

抽取“法律角色”下的“被告人”子类，抽取被告人姓名，并标记为“@被告人姓名”
抽取性别、年龄、出生年月、出生地、民族、住址、职业、户籍地、涉嫌罪名、被羁押日期、被逮捕日期、现羁押地的信息，并标记}
if 匹配“辩护人”子类 {
抽取“法律角色”下的“辩护人”子类，抽取辩护人姓名，并标记为“@辩护人姓名”
抽取辩护人单位等信息，并标记}
if 匹配“法定代理人”子类 {
抽取“法律角色”下的“法定代理人”子类，抽取法定代理人姓名，并标记为“@法定代理人姓名”
抽取法定代理人年龄、出生日期、民族、职业、住址、与被告关系等信息，并标记}
}
.....
标注关系：被告人——辩护人
标注关系：被告人——代理人
}

对判决书中的法律角色、起诉书、案情、证据、判决 5 个子类进行抽取标注，表 3-3 和表 3-4 为判决书和案情的抽取模式，其他部分略。

表 3-3 抽取模式——判决书 5 个子类

XML 标记	类 名	模 式	类 型
Roles	法律 角色	Pattern1=被告人.*\n Pattern2=(法定)?代理人.*\n Pattern3=辩护人.*\n	Multiline
Charge	起诉书	Pattern1=[\u4E00-\u9FFF]+院以 [\u4E00-\u9FFF]+[\d{4}][\d{3,6}]号.*\r Pattern2=. *院(起诉书)?指控, .*\r Pattern3=针对(上述)?指控(的)?(上述)?(事实)?, 公诉(机 关 人)(当庭 向法庭).*\r Pattern4=. *(((被告人[\u4E00-\u9FFF]{2,4}的[其]辩护人 (的辩护意见是 认为)))(被告人[\u4E00-\u9FFF]{2,4}及辩护 人均认为.*))	Multiline
Case	案情	经(本院)?审理查明, .*\r 被告人.*(归案 投案)	Multiline

续表

XML 标记	类 名	模 式	类 型
Proof	证据	(?<=本院予以确认的下列证据证实 上述事实，有下列证据证实：)(.*r)+(?=本院认为)	Multiline
Judge	判决	本院认为，(.r)+(?=审判长)	Multiline

表 3-4 抽取模式——案情

XML 标识	类 名	模 式
Case_time	案发时间	\d{4}年\d{1,2}月\d{1,2}日(晚\d{1,2}时许)?
Case_place	案发地点	(在 位于).*(附近 外 内 侧)
Case_desendent_name	案情_被告人姓名	结合实体抽取部分识别的被告人
Case_victim_name	案情_被害人姓名	对案情模块进行句法分析后，结合情节描述的上下文识别被害人
Case_victim_number	案情_被害人个数	Count（被害人姓名）
Case_affair	事件	因[\u4E00-\u9FFF]+(?=。 ，)，需结合句法分析
Case_violenceway	施暴方式	厮打 互殴 殴打 猛刺 击打，需结合句法分析，抽取类似意义的动词部分
Case_attackpart	攻击部位	((胸(腹)?部) 腹部 (头(面)?部) 四肢部 背部 肋部 颈项部 (左 右)?肩部 身体)(等)?
Case_attacktimes	攻击次数	抽取相关数词或模糊数词，如“十余刀或多次或几次”
Case_attacktool	作案工具	((?<=持)([\u4E00-\u9FFF]+、?)+(?=猛刺 朝 等)) (?<=被)[\u4E00-\u9FFF]+(?=(多次)?击打)
Case_result	伤情等级	致.*(死亡 (轻(微)?伤) 重伤)

各类实体及其属性的抽取模式的集合，利用工具软件构造模式抽取器的 XML 模板，用 XML 语言表示出来，形成该罪名的抽取标注模板，存入到本体知识库。以故意伤害罪为例，其基于本体的 XML 抽取模板见附录 D。

3. 刑事案件案例标注抽取过程

刑事案件信息抽取过程，概括起来讲是经过 4 个环节，抽取标注 3 类信息。4 个环节，也可以称作 4 个模块，即自然语言处理模块、模式识别模块、信息

抽取模块及格式转化模块，各模块相互配合进行横向的信息内容的抽取。抽取标注的 3 类信息是命名实体、实体属性及实体关系。需要结合文本挖掘工具进行文本分析，结合正则表达式进行规则建立及结合 C# 编程进行抽取和格式转化。

由于相同罪名的案例文档集是一系列具有相似结构的文档，但同时又包含自然语言的描述，信息的抽取和标注需要结合正则表达式对文本中的实体和规范模式的部分进行信息抽取，对不能直接抽取的部分需要结合自然语言处理进行信息抽取。抽取流程如图 3-11 所示。

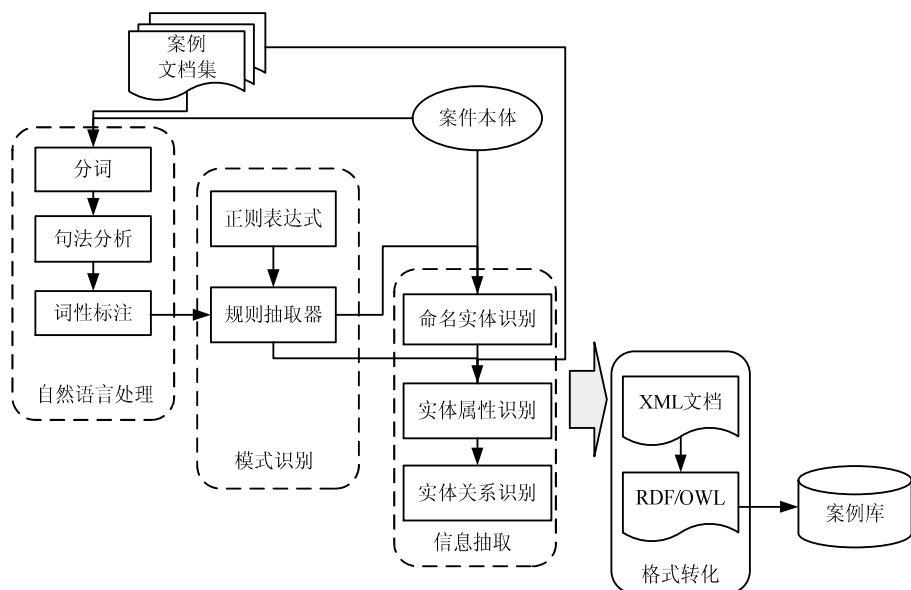


图 3-11 信息抽取与标注过程

首先运用 ICTCLAS 软件结合案件本体构造分词词典，对案例文档集进行分词，然后进行句法分析和词性标注，运用正则表达式建立 XML 文档形式的规则抽取器（见附录 D），结合规则抽取器和上下文信息进行命名实体的识别。这里的命名实体主要是指包含被告人、辩护人等案件角色的信息。因为这些信息会在文档中多次提及，但有些情况下未给出角色标注，所以需要先对这些信息

进行抽取,以便在实体属性抽取时进行识别。例如,对于“被告人/n×××/n”,其中“被告人”为本体中的类名,则可以依据正则表达式“(?!<被告人/n)[\u4E00-\u9FFF]+(?!/n)”对被告人后的姓名进行抽取,并在×××左右添加<Role_defendent_name>和</Role_defendent_name>。

其次是对实体属性的识别。主要是运用正则表达式构造的规则抽取器,按照案例文档结构对文档内容进行分块,对应本体构造6个子类,判决书基本信息、法律角色、起诉书、案情、证据和判决。然后结合之前抽取出的本体实例对各类的详细属性进行抽取,如“×市人民检察院第二分院以京检二分刑诉[2007]183号起诉书指控被告人×××犯故意伤害罪”中,可以依据模式“(?!<以)[\u4E00-\u9FFF]+([\d{4}][\d{3,6}]、?)+号”抽取出起诉书编号。

第三是对抽取出的信息之间的对应及从属关系进行抽取。如合议庭与法院的从属关系,检察员与检察院的从属关系,被告人与代理人的对应,被告人与辩护人的对应关系等。对以上抽取的内容添加XML标记,并存入XML模板。

最终对抽取结果进行汇总,对XML结构进行整理,建立对应罪名的标注库。再将标注库中的信息转换成RDF或OWL的形式,便于向本体库中导入,形成最终的案件本体实例库,即案例库。

3.2.4 刑事案例库设计

刑事案件案例库的设计要满足决策支持系统的要求,决策支持系统数据库所要求的数据要具备三方面特点:概括性、抽象性、统一性。刑事案例数据库设计的关键是数据库的结构设计,包括逻辑设计和物理设计。

1. 案例库逻辑设计

刑事案件本体是刑事案件的概念模型，它是整个刑事案件的信息结构，逻辑设计的目的是将这一刑事领域模型转换为与数据库系统所支撑的数据模型相符合的逻辑结构。由于本书采用的大多是关系型数据库，所以逻辑设计的主要问题就是把各个实体与实体属性的联系转换为关系模式。综合上述研究，表 3-5 列出了刑事案件本体概念模型与案例库关系模型的对应关系。

表 3-5 概念模型与案例库逻辑设计关系

环 节	对 象	对象属性	属 性 值
刑事案件本体	类/子类	属性（数据、对象属性）	
案例集抽取标注	实体	属性（属性集、关系）	文字描述
案例库	主表/子表	字段（主表到子表的主键）	规范字段值

从表 3-5 中可以看出，案例库的逻辑设计是将刑事案件作为主表，将刑事案件本体中的类/子类作为子表，将刑事案件本体的属性作为研究的字段，通过主键建立起主表与子表的附属关系，将从案例文档集中抽取的属性和属性值作为案例库表中字段的内容。需要说明的是，这里的字段值是在案件信息抽取标注得到的文字描述信息基础上，根据内容特点进行转化来表示的，这样便于数据挖掘利用。

针对案例文档集抽取标注得到的文字描述信息和实体属性的特征，将属性值主要分为数值型（如判决刑期）、字符型（如审判员名称）和其他几种。研究中发现，为便于对数据库的应用，对于是否存在某种情况的属性用 0/1 表示，对于等级数据用字符表示，这种转化使得数据库更便于数据分析。表 3-6 以故意伤害罪中的伤情等级和故意伤害罪的法规表示为例进行说明，在案例库中，根据数据字典，存储“编号”信息。在第 5 章的推理模型建立中，就是以此案例库为基础的。

表 3-6 属性值规范化表示

量刑类别	编 号	伤情等级	编 号
三年以下有期徒刑	E1	轻微伤	D1
三年以下拘役	E2	轻伤（下限）	D2
三年以下管制	E3	轻伤（偏轻）	D3
三年以上十年以下有期徒刑	E4	轻伤	D4
十年以上有期徒刑	E5	轻伤（偏重）	D5
无期徒刑	E6	轻伤（上限）	D6
死刑	E7	重伤（下限）	D7

2. 案例库的物理设计

数据库在物理设备上的存储结构与存取方法称为数据库的物理结构。好的物理结构设计能使数据库上运行的各种事务存储空间利用率高、响应时间小、事务吞吐率大。物理设计中最重要的是有一个高效率的存取方法，常见的存取方法有索引存取方法和 HASH 存取方法等。本书采用的是 Sybase IQ 数据仓库，对数据采用列式存储，实现了无处不索引的目标，使数据存取高效。但这不是本书研究的重点，这里不再详细叙述。

3. 刑事案件知识的数据化表示

按照支持量刑决策分析、审判指标时间序列规律分析、审判动态数据地图分布分析和审判数据关联分析的目的，刑事案件信息以案例库与规则库、模型库、分析库、知识库等一起共同实现刑事案件知识的表示。

案例库中存储的是从判决书中抽取并格式转化之后，用于支持量刑决策的历史案件信息的结构化数据，如附录 A 以故意伤害罪为例构建的刑事案件本体类与属性的详细设计中的内容，以及附录 E 中的故意伤害罪的 39 个主要特征。

此外,还有以实体内容存储的刑事案件判决书,它是一个不断增加和调整的库,根据决策目标和新案例的加入,按一定时间和策略进行调整。

规则库中存储的是将我国的《刑法》中的条款,根据量刑决策需要进行细化的结果,目的是将量刑区间进一步细化,将成文法典中的模糊词进行量化,以便在量刑决策中使用,如表 3-6 中的内容。

模型库中存储的是针对不同刑事案件类刑(罪名),而形成的针对该案件类型进行信息抽取的刑事案件本体 XML 抽取模板、犯罪特征向量的各维度权重配置、两种决策推理模型算法等。这样可以在推理时,根据目标案例类型,调取相应的案件本体抽取模板和推理集成模型。

分析库是指经过日常积累、整理形成的,反映量刑决策特质的、指标的现象的集合,实现一张表反映一个事件(如有相似事件可叠加)。例如,刑事案件中的从轻、从重情节的适用问题。从轻、从重情节的适用在法条中定义较为宽泛,而在实际审判、案件调查过程中从轻、从重情节存在具体表现。因现象之间可进行互相比照,故分析库通过大量历史案件情节的归纳、整理,提取出法官认定的从轻、从重情节作为表象,为法官审判工作提供参考。

知识库存储的是司法审判领域已经发现或形成的审判规律,以及通过(文本)数据挖掘新发现的审判规律。例如,优秀案例或参阅案例的判定、法院年人均结案数的范围、可以接受的一审陪审率、陪审员参加诉讼对案件公正的作用力、专业法律人员(律师等)对保护当事人权益的价值等。

以上,以案例库为核心,共同为量刑决策等司法审判规律的探索提供支撑。

3.3 刑事案件本体构建实现与信息抽取评价

本节对刑事案件本体构建进行实证研究，同时对利用本体方法进行刑事案件信息抽取结果进行评价。

3.3.1 刑事案件本体构建实现

1. 本体构建实现方法

使用本体编辑工具 Protégé 4.1，以故意伤害罪为例构建故意伤害罪类的刑事案件本体，并生成了以 OWL 描述的形式化本体模型。

2. 本体构建实例

首先，在 Class Hierarchy 选项卡中编辑类的层级结构，将判决书、判决结果、合议庭、案情、法律机构、法律角色、法条、证据、起诉书作为 Protégé 顶层类 Thing 的子类，如图 3-12 所示。

其次，进行本体属性设计，即在 Data Properties 选项卡的 Data Property Hierarchy 窗口编辑类的数据属性和对象属性，如图 3-13 所示。

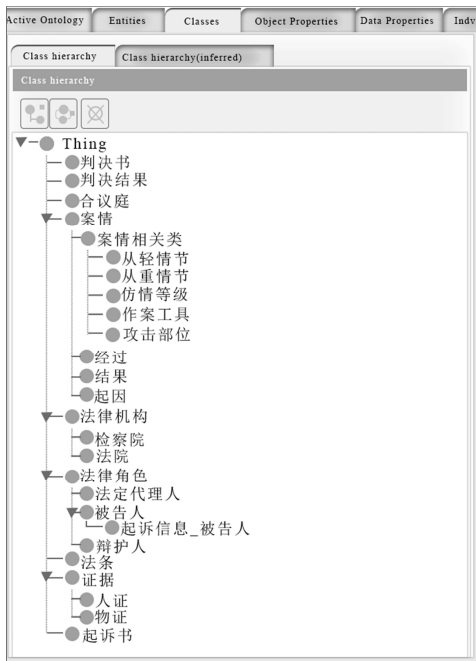


图 3-12 刑事案件本体类架构

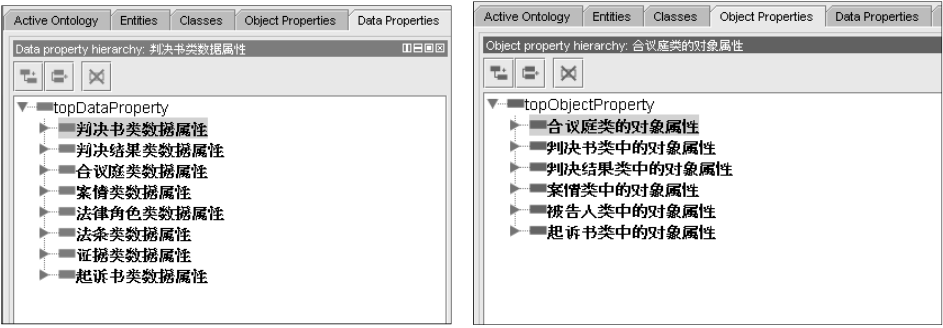


图 3-13 数据属性架构

最后，实现刑事案件本体实例。这里仅以刑事案件本体中的“案情类”为例，介绍在刑事案件本体中实例的组织方式。

在 Protégé 的 Individuals 选项卡中添加用于表示起诉信息中案情的案情类实例案情_京检二分公诉二刑诉[2005]×××号和用于表述判决书中审理查明的案情类实例案情_2006 二中刑初字第×××号，并为实例添加对象属性和数据属性并根据案例文本为属性赋值，如图 3-14 所示。然后依次建立起诉书类实例——京检二分公诉二刑诉[2005]×××号和判决书类实例——2006 二中刑初字第×××号，为其添加对象属性和数据属性并根据案例文本为属性赋值。



图 3-14 a) 案情类实例示例



图 3-14 b) 案情类实例示例所包含的属性和属性值

图 3-15 所示为使用 Protégé 插件 OntoGraf 所绘制的判决书类、起诉书类、案情类实例的组织结构。图中的实线 D、E、F、G 表示类-实例的关系；图中上半部分的虚线 A、B、C 分别表示类之间的对象关系 Has 起诉信息、Has 起诉信息_案情、Has 案情；图下半部分的虚线 H、I、J 表示为实例赋值的对象关系。

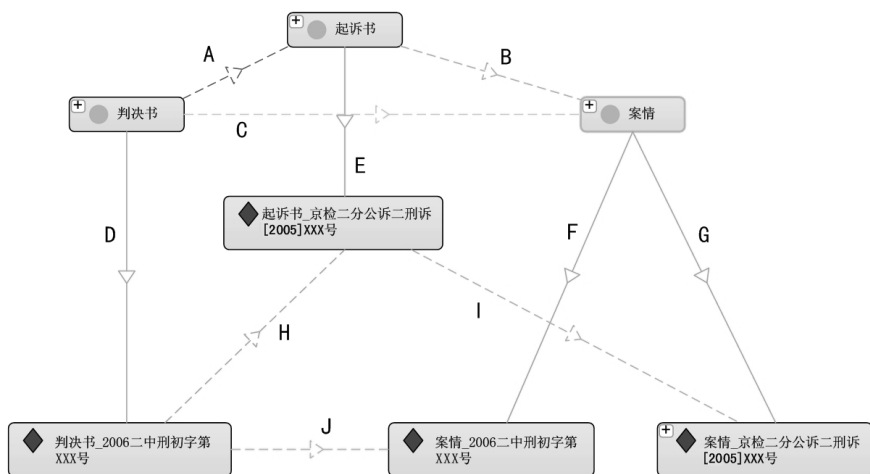


图 3-15 本体实例组织示例

除了 3.2.2 中图示的本体模型外，还可以用描述性刑事案件本体模型。下面就是按 $O_{\text{刑事案件}} = \langle T, X, XD, AA, TR \rangle$ ，其中， $T = \langle TC, TD, TO \rangle$ 得到的故意伤害罪类的刑事案件本体的部分内容，详细结果见附录 C。

$TC = \{ \text{ 法律机构, 法院, 检察院, 起诉书, 合议庭, 法律角色, 被告人, 起诉信息_被告人, 辩护人, 法定代理人, 案情, 起因, 经过, 结果, 案情相关类, 攻击部位, 作案工具, 伤情等级, 从轻情节, 从重情节, 证据, 人证, 物证, 判决结果, 法条, 判决书 } \}$

3.3.2 刑事案件信息抽取评价

1. 实验数据与方法

采用北京某法院 2006—2010 年的 1 000 个故意伤害罪的判决书为实验数据 (文档长度在 5~15KB), 以故意伤害罪类 XML 抽取模板为工具 (附录 D), 对信息抽取标注模型进行评价。实验平台采用 Visual Studio 2008 并结合 C# 编程语言。

这里需要特别说明的是, 在本研究中, 选取了 11 000 个故意伤害罪判决书, 创建了故意伤害罪刑事案件本体, 以此形成抽取模板 (见附录 D), 对 11 000 个案例实现了自动抽取。本研究对此评价实验中, 选择了其中的 1 000 个判决书, 采用人工核对的方法对信息抽取的是否准确进行判断, 以此判断信息抽取的准确率, 因为信息抽取的准确率只能采用主观的判断进行确定; 同时, 根据每组抽取结果和新发现的半结构化文本中的数据特征, 对抽取模板进行优化。

2. 实验与评价

根据 3.2.3 中制定的抽取策略, 以故意伤害罪的刑事案件本体 XML 抽取模板和刑事判决书文档集为输入, 对故意伤害罪的 85 项重要信息进行抽取, 对每一个判决书抽取标注后的 XML 文档如表 3-7 所示, 将抽取后结果与人工核对结果进行比较, 用抽取正确的信息数据占 85 项信息的比率作为信息抽取的准确率。

表 3-7 抽取标注后对应的 XML 文档部分

XML 文档
<pre><?xml version="1.0" encoding="UTF-8"?> <Judgement> <Entities_JudgementCName="刑事判决书">判决书_（2008）二中刑初字第×××××号 <Roles CName="法律角色"> <Role_defendentCName="被告人">张×× <Role_defendent_nameCName="被告人姓名">张×× </Role_defendent_name> <Role_defendent_sexCName="被告人性别">男 </Role_defendent_sex> <Role_defendent_ageCName="被告人年龄">46 岁 </Role_defendent_age> <Role_defendent_birthdateCName="被告人出生年月">1962 年 5 月 16 日 </Role_defendent_birthdate> <Role_defendent_birthplaceCName="被告人出生地">山西省孟县 </Role_defendent_birthplace> <Role_defendent_nationCName="被告民族">汉族 </Role_defendent_nation> <Role_defendent_educationCName="被告人文化水平">小学文化 </Role_defendent_education> <Role_defendent_careerCName="被告人职业">×市×××公司职工 </Role_defendent_career> <Role_defendent_addressCName="被告人住址">×市朝阳区慈云寺商场院内平房 </Role_defendent_address> <Role_defendent_residenceCName="被告人户籍地">×市朝阳区关东店×巷××号 </Role_defendent_residence> <Role_defendent_crimeCName="涉嫌罪名">故意伤害罪 </Role_defendent_crime> <Role_defendent_detainedCName="被羁押日期">2008 年 3 月 17 日 </Role_defendent_detained> <Role_defendent_arrestedCName="被逮捕日期">同年 3 月 28 日 </Role_defendentCName="被告人"> </Roles CName="法律角色"> </Entities_JudgementCName="刑事判决书"> </Judgement></pre>

续表

XML 文档
<div></Role_defendent_arrested></div> <div><Role_defendent_detainplaceCName="现羁押地">×市第一看守所</div> <div></Role_defendent_detainplace></div> <div></Role_defendent></div> <div><Role_defenderCName="辩护人">王建京</div> <div><Role_defender_nameCName="辩护人姓名">王建京</div> <div></Role_defender_name></div> <div><Role_defender_unitCName="辩护人单位">×市京工律师事务所律师</div> <div></Role_defender_unit></div> <div></Role_defender></div> <div>.....</div> <div></Entities_Judgement></div> <div><Relations_Judgement></div> <div><Relation_defendRName="Has 辩护人"></div> <div><Source CName=“被告人”>张××</Source></div> <div><Domain CName="辩护人">王建京</Domain></div> <div></Reltion_Defend></div> <div>.....</div> <div></Relations_Judgement></div> <div></Judgement></div>

实验将 1 000 个案例文档分 10 组进行实验，采用召回率和准确率两个指标作为模型的评价标准，实验结果如表 3-8 所示。在实验中，依据抽取结果中出现的问题对抽取模型进行逐步修正，最终使抽取结果的召回率和准确率都达到 80% 以上。

表 3-8 实验结果分析

组 别	1	2	3	4	5	6	7	8	9	10
召回率%	65.2	68.1	71.2	78.4	83.9	87.8	86.3	89.2	88.7	89.4
准确率%	52.3	57.8	63.5	69.4	72.6	73.9	78.1	80.2	82.6	81.4

经过实验和对没有实现 100%信息抽取的案例分析发现：没有抽取到的信息主要是法官撰写文书的格式及一些特殊字符造成的，这些信息不影响后续决策推理对基础信息的要求。因此，该模型的抽取结果达到预期效果，证明了该方法将半结构化决策数据转化为规则化案例库的有效性。

3.4 本章小结

本章从司法审判活动对案例需求和数据挖掘机制出发，首先，以刑事案件的案例文档集为研究对象，通过分析法律文书的特征，提出了从领域本体模型构建到信息抽取标注，再到案例库设计的一体化文书规范处理模式。其次，依据本体建立基本理论，提出了刑事案件本体的设计模型，以案件类型（罪名）为单位，建立相应的刑事案件本体，形成刑事案件本体库，对刑事案件中相关的概念及其关系进行了全面梳理，提出了在司法审判领域进行知识表达的新方法。第三，基于刑事案件本体对刑事案件全貌的知识化整理，结合信息技术，实现了对案例文档集的抽取与标注，实现了实体信息从类别到属性的描述性抽取，建立了抽取标注的通用模板。第四，按照数据仓库方便存储、快速检索，支持数据分析的要求，对案例库进行了逻辑设计和物理设计，完成了从本体、抽取信息到可供数据挖掘存储数据的转换，实现了从“沉睡的”海量数据向支持决策的“有价值的”信息的转化。第五，对本章提出的本体构建理论及信息抽取标注方法进行了验证和评价，结果表明：刑事案件本体表示具有一致性；信息抽取标注方法准确率达到 80% 以上，能够满足决策推理对案例库的基础数据要求。

基于遗传算法的刑事案例属性优化 与检索策略研究

半结构化判决书文档集结构化后，形成了反应刑事案件全部信息的案例库，它具有案例多、属性多、要素多的特点，直接用于支持法官决策会呈现检索速度慢、效率低，达不到有效支持量刑决策的目的。本章研究问题是针对多案例、多属性、多要素案例库，如何选择和确立案例属性优化和案例快速检索机制，解决法官量刑决策的关键技术问题。研究思路是以遗传算法为核心，结合 KNN（K-Nearest Neighbor, KNN）方法，论证案例属性优化与检索机制，通过对比论证其作为案例检索方法的优势，最终构建满足训练决策推理模型要求的案例检索方法与相应案例库。

4.1 基于遗传算法对刑事案例属性优化的研究

目前的 CBR 分析中,案例检索有多种方法,其中最常用的有知识导引法、神经网络法、KNN 方法、灰色关联方法等。从目前文献中所阐述的各种方法的应用环境、方法的准确性与适应性看,各种方法在不同的背景下显出不同的应用精度,KNN 算法因其工作原理简单、精度高而应用较广。但这一算法在实际应用中有两个难以逾越的障碍,即案例特征权重确立与检索时间随着案例数增加呈线性增长,在海量数据下,影响 KNN 方法的应用效率。对于一些多案例,这一算法的应答时间会大幅度加长。因此,方法的选择要依据一定的应用环境。

4.1.1 刑事案例属性的特征及属性优化的意义

第3章以刑事案件的案例文档集为研究对象,归纳总结出以文档集为决策依据的领域,通过本体建立、信息抽取到案例库构建,使决策信息结构化、规则化,按照司法案例的特征建立了基于本体的可挖掘的案例库。与一般案例库相比,司法审判类案例库有其特殊性,案例属性维度、案例标准化难度、案例定量性、案例属性值的多样性、案例参数的类型等与其他类型的案例有较大差异。图4-1的雷达图是对期刊网中的200篇案例推理文献进行的比较分析,选用五个属性对各类不同案例进行比较,从中可以看出刑事案例与其他类型案例的不同,明显具有多维度、标准化难、定量难、属性值多样和案例参数类型多的复杂性特点。

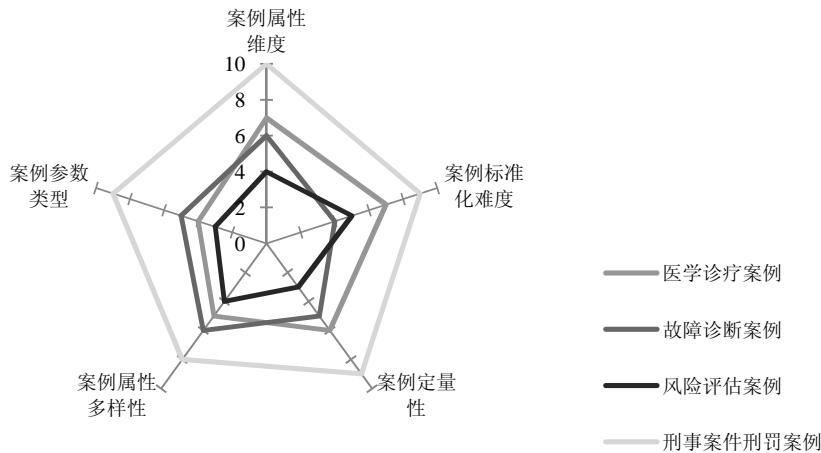


图 4-1 不同类型案例特征比较

因此，选择科学的、有针对性的方法形成对案例属性优化和案例的有效检索是非常必要的。哪种方法对刑事案例的检索及属性优化有较好的优势，本书将对常用的方法进行对比、论证，选择恰当的属性优化方法。采用遗传算法和改进的层次分析法（Analytic Hierarchy Process, AHP）对案例属性进行优化，从而形成适合司法刑事案例的属性优化方法。

4.1.2 两种优化方法对刑事案例属性优化的针对性分析

一般的特征权重确定方法有两大类型，一是主观确权法，即由选定的专家根据行业经验判断各属性相对重要性程度进行指标权重确定；二是客观确权法，即各属性指标根据一定规则进行自动确定权重的方法，如关联分析法、回归分析法等。运用主观确权法确定各因素指标的权重值反映了决策者的意向，但结果的主观随意性较大。客观确权法虽然受主观人为因素影响较小，确定的权重更精确，但仍然受原始数据的限制，也不能广泛应用到实际中。所以，CBR 分

析中的方法,应通过不同类型方法比较来确定。本书采用代表主观确权方法的改进 AHP 方法、代表客观确权方法的遗传算法进行比较分析。

1. 基于遗传算法对案例属性的优化分析

遗传算法 (Genetic Algorithm, GA) 是由美国 J.Holland 教授 1975 年提出,随之形成一种应用性很广的随机化搜索方法。这种搜索方法是借鉴生物界的进化规律演化而来的一种优化方法,其特点在于以概率化寻优方法获取和指导优化的搜索空间,自适应地调整搜索方向,形成优化的结果。

遗传算法兼有生物进化和遗传的主要思想,在组合优化方面这一方法具有与传统优化方法所不同的特征。一是遗传算法的处理对象集中于问题参数的编码集,这种编码集运算简单、可以表达复杂问题。对于许多复杂系统问题,遗传算法因其适应条件宽泛、不受多种条件约束而得到很好的应用。二是遗传算法能对搜索空间中多个点进行求解寻优,从而减少了解收敛于局部最小的可能,这样便增加了处理问题的并行性,避免了局部最优。三是遗传算法一般利用随机性规则来引导搜索,从而具有较强的搜索能力。四是这种算法可以很好地同其他技术进行结合使用,因此可扩展性较强。可见,遗传算法由于受约束少、可随机搜索而具备很好的搜索效率。

由于遗传算法直接对研究的结构对象进行针对性操作,不存在一定连续性的限定,因而具有内在的隐并行性和更好的全局寻优能力。采用概率化的寻优方法,能自动获取和指导优化的搜索空间,自适应地调整搜索方向,不需要确定的规则。这些特质已经被广泛地应用于组合优化、自适应控制和人工生命等领域。因此,运用这一方法进行属性优化与检索,对司法审判类型的案例检索有很好的针对性、适用性。基于对刑事案例库的研究,本书采用遗传算法进行案例属性权重确定与案例检索。

2. 基于改进 AHP 方法对案例属性的优化分析

传统 AHP 算法是通过专家对属性间关系进行主观判断, 从而形成相应判断矩阵, 进而实现属性优化分析。这一方法需要进行一致性判断, 并且对判断矩阵进行调整。对于本书所研究的刑事案例, 体现出属性多、要素多、案例类型多等特点, 如果一致性判断未能达到要求, 重新调整有较大的难度, 在研究中如果主观判断多, 对评价结果会产生较大的影响。为克服这一缺陷, 在优化分析中我们采用改进的 AHP 方法。这一方法将判断矩阵的标度简化为三阶度, 同时不需要进行一致性判断, 从而使研究精度增加, 更适合司法案例研究的特征。这一方法减少了判断指标重要程度时出现的一定程度的模糊性; 此外, 通过引入最优传递矩阵, 对矩阵一致性有一定保证。

改进的 AHP 方法研究步骤如下。

首先, 根据案例属性之间的重要性, 分析特征属性的权值, 如判断矩阵式为

$$A = \begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{n1} & \cdots & a_{nn} \end{bmatrix} \quad (4-1)$$

其中,

$$a_{ij} = \begin{cases} 1, & i \text{ 比 } j \text{ 重要} \\ 0, & i \text{ 与 } j \text{ 同等重要} \\ -1, & j \text{ 比 } i \text{ 重要} \end{cases} \quad (4-2)$$

判断矩阵 A 的最优传递矩阵 B 可由式 (4-3) 得出。

$$b_{ij} = \frac{1}{n} \sum_{k=1}^n (a_{ik} + a_{kj}) \quad (4-3)$$

最优传递矩阵可转换为一致性矩阵 C , 其中的元素 c_{ij} 如式 (4-4)。

$$c_{ij} = e^{b_{ij}} \quad (4-4)$$

通过式(4-4)可计算出矩阵 C 的特征值,最大特征值的特征向量对应特征的权重向量。

综上所述,从案例的模糊性、案例属性维度、案例标准化、案例的定量难度、案例属性值的多样性、案例参数的类型等表明刑事案例与其他类型的案例有较大差异,而权重的确定有重要意义。从遗传算法、改进 AHP 方法的特征看,均可以作为案例属性权重的确定方法,以下可以从两种方法的比较过程找到适合这一研究的科学方法。

4.1.3 针对刑事案例属性优化的两种优化方法对比分析

1. 基于改进的 AHP 方法对案例属性优化分析结果

由于刑事案例属性数量多、类型各异、定量化分析难度大,在案例属性优化过程中的许多方法都难以适用。而改进的 AHP 方法将九维度的判断改进为三维度,从而不需要经过一致性判断检验,经过建立比较矩阵、重要性排序、构造判断矩阵、分析优化传递矩阵等一系列过程,就能得到合适的刑事案例属性优化结果。这一方法可以作为刑事案例属性优化的备选方法之一。

改进的 AHP 方法通过七步实现属性优化:① 建立 AHP 模型;② 建立比较矩阵;③ 对矩阵进行重要性排序;④ 构造判断矩阵;⑤ 求判断矩阵 B 的最优传递矩阵 C ;⑥ 求拟优一致矩阵 D ;⑦ 求 D 的最大特征值对应的特征向量,并进行归一化处理,进而得各因素权重。

(1) 应用改进的 AHP 模型求解刑事案例属性权值。以故意伤害罪为例,建立改进的 AHP 模型,如图 4-2 所示。

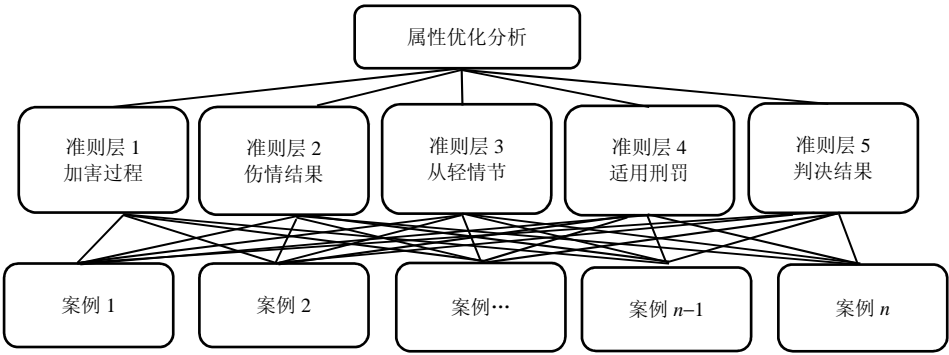


图 4-2 改进的 AHP 结构图

(2) 建立比较矩阵。用神经网络方法对 1 000 个刑事案例，对故意伤害罪中的情节与判决结果相关性进行判断，确定出施暴方式、作案工具、袭击部位、加害结果、适用法律和从轻表现 6 个方面 25 个案件判断准则层指标，见表 4-1。经 20 名刑事法官进行属性重要性判断，得出判断矩阵 E 。

$$E = \begin{bmatrix} e_{11} & \cdots & e_{1n} \\ \vdots & \ddots & \vdots \\ e_{n1} & \cdots & e_{nn} \end{bmatrix} \tag{4-5}$$

其中

$$e_{ij} = \begin{cases} 2, & \text{第 } i \text{ 个因素比第 } j \text{ 个因素重要} \\ 1, & \text{第 } i \text{ 个因素与第 } j \text{ 个因素同等重要} \\ 0, & \text{第 } i \text{ 个因素没有第 } j \text{ 个因素重要} \end{cases} \tag{4-6}$$

表 4-1 案例属性

序 号	犯罪情节	属性名称	属性具体内容
1	施暴方式	M1	打
		M2	刺
		M3	其他
		M4	人数

续表

序 号	犯罪情节	属性名称	属性具体内容
2	使用工具	M5	拳脚
		M6	钝器
		M7	锐器
3	袭击部位	M8	胸部
		M9	腹部
		E10	头部
		E11	四肢
		E12	其他
4	加害结果	E13	D4
		E14	D6
		E15	D8
		E16	D11
5	从轻环节	M17	自首
		M18	积极救治
		M19	认罪
		M20	悔罪
		M21	积极赔偿
6	刑罚类型	M22	刑罚 E1
		M23	刑罚 E4
		M24	刑罚 E5
		M25	刑罚 E6

按改进的 AHP 方法步骤，按重要性程度重新排列，从而得出比较矩阵，以下为 25 个案件判断准则层得出的比较矩阵。

1	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2
0	1	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2
0	0	1	1	1	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2
0	0	0	1	1	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2
0	0	0	0	1	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2
0	0	0	0	0	1	1	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2
0	0	0	0	0	0	1	1	1	1	1	1	1	1	1	2	2	2	2	2	2	2	2
0	0	0	0	0	0	0	1	1	1	1	1	1	1	1	1	1	1	2	2	2	2	2
0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1	1	1	2	2	2	2	2
0	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1	1	2	2	2	2	2
0	0	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1	2	2	2	2	2
0	0	0	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	2	2	2	2	2
0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1	1	1	2	2	2	2	2
0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1	1	2	2	2	2	2
0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1	2	2	2	2	2
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	2	2	2	2	2
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	2	2	2	2	2
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	2	2	2	2	2
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1	1
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1	1
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1	1
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1	1
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1	1
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1	1

(3) 由式 (4-7) 确定矩阵重要性排序指数 r_i 。

$$r_i = \sum_{j=1}^n e_{ij} \quad (4-7)$$

可知: $r_{\max} = \max(r_i)$ $r_{\min} = \min(r_i)$

(4) 构造判断矩阵, 按照式(4-8)和式(4-9)得出判断矩阵 F , 如图 4-3 所示。

$$f_{ij} = \begin{cases} \frac{r_i - r_j}{r_{\max} - r_{\min}}(k_m - 1) + 1, r_i \geq r_j \\ \left[\frac{|r_i - r_j|}{r_{\max} - r_{\min}}(k_m - 1) + 1 \right]^{-1}, r_i < r_j \end{cases} \quad (4-8)$$

$$k_m = \frac{r_{\max}}{r_{\min}} \quad (4-9)$$

1	1.33	2	2.167	2.33	2.83	4.33	5.17	5.33	5.5	5.67	5.67	6	6.167	6.33	6.5	6.667	6.833	7	8.167	8.167	8.17	8.167	8.167	8.167
0.75	1	1.67	1.833	2	2.5	4	4.83	5	5.17	5.33	5.33	5.67	5.833	6	6.17	6.333	6.5	6.667	7.833	7.833	7.83	7.833	7.833	7.833
0.5	0.6	1	1.167	1.33	1.83	3.33	4.17	4.33	4.5	4.67	4.67	5	5.167	5.33	5.5	5.667	5.833	6	7.167	7.167	7.17	7.167	7.167	7.167
0.462	0.55	0.86	1	1.17	1.67	3.17	4	4.17	4.33	4.5	4.5	4.83	5	5.17	5.33	5.5	5.667	5.833	7	7	7	7	7	7
0.429	0.5	0.75	0.857	1	1.5	3	3.83	4	4.17	4.33	4.33	4.67	4.833	5	5.17	5.333	5.5	5.667	6.833	6.833	6.83	6.833	6.833	6.833
0.353	0.4	0.55	0.6	0.67	1	2.5	3.33	3.5	3.67	3.83	3.83	4.17	4.333	4.5	4.67	4.833	5	5.167	6.333	6.333	6.33	6.333	6.333	6.333
0.231	0.25	0.3	0.316	0.33	0.4	1	1.83	2	2.17	2.33	2.33	2.67	2.833	3	3.17	3.333	3.5	3.667	4.833	4.833	4.83	4.833	4.833	4.833
0.194	0.21	0.24	0.25	0.26	0.3	0.55	1	1.17	1.33	1.5	1.5	1.83	2	2.17	2.33	2.5	2.667	2.833	4	4	4	4	4	4
0.187	0.2	0.23	0.24	0.25	0.29	0.5	0.86	1	1.17	1.33	1.33	1.67	1.833	2	2.17	2.333	2.5	2.667	3.833	3.833	3.83	3.833	3.833	3.833
0.182	0.19	0.22	0.231	0.24	0.27	0.46	0.75	0.86	1	1.17	1.17	1.5	1.667	1.83	2	2.167	2.333	2.5	3.667	3.667	3.67	3.667	3.667	3.667
0.176	0.19	0.21	0.222	0.23	0.26	0.43	0.67	0.75	0.86	1	1	1.33	1.5	1.67	1.83	2	2.167	2.333	3.5	3.5	3.5	3.5	3.5	3.5
0.176	0.19	0.21	0.222	0.23	0.26	0.43	0.67	0.75	0.86	1	1	1.33	1.5	1.67	1.83	2	2.167	2.333	3.5	3.5	3.5	3.5	3.5	3.5
0.167	0.18	0.2	0.207	0.21	0.24	0.37	0.55	0.6	0.67	0.75	0.75	1	1.167	1.33	1.5	1.667	1.833	2	3.167	3.167	3.17	3.167	3.167	3.167
0.162	0.17	0.19	0.2	0.21	0.23	0.35	0.5	0.55	0.6	0.67	0.67	0.86	1	1.17	1.33	1.5	1.667	1.833	3	3	3	3	3	3
0.158	0.17	0.19	0.194	0.2	0.22	0.33	0.46	0.5	0.55	0.6	0.6	0.75	0.857	1	1.17	1.333	1.5	1.667	2.833	2.833	2.83	2.833	2.833	2.833
0.154	0.16	0.18	0.187	0.19	0.21	0.32	0.43	0.46	0.5	0.55	0.55	0.67	0.75	0.86	1	1.167	1.333	1.5	2.667	2.667	2.67	2.667	2.667	2.667
0.15	0.16	0.18	0.182	0.19	0.21	0.3	0.4	0.43	0.46	0.5	0.5	0.6	0.667	0.75	0.86	1	1.167	1.333	2.5	2.5	2.5	2.5	2.5	2.5
0.146	0.15	0.17	0.176	0.18	0.2	0.29	0.37	0.4	0.43	0.46	0.46	0.55	0.6	0.67	0.75	0.857	1	1.167	2.333	2.333	2.33	2.333	2.333	2.333
0.143	0.15	0.17	0.171	0.18	0.19	0.27	0.35	0.37	0.4	0.43	0.43	0.5	0.545	0.6	0.67	0.75	0.857	1	2.167	2.167	2.17	2.167	2.167	2.167
0.122	0.13	0.14	0.143	0.15	0.16	0.21	0.25	0.26	0.27	0.29	0.29	0.32	0.333	0.35	0.37	0.4	0.429	0.462	1	1	1	1	1	1
0.122	0.13	0.14	0.143	0.15	0.16	0.21	0.25	0.26	0.27	0.29	0.29	0.32	0.333	0.35	0.37	0.4	0.429	0.462	1	1	1	1	1	1
0.122	0.13	0.14	0.143	0.15	0.16	0.21	0.25	0.26	0.27	0.29	0.29	0.32	0.333	0.35	0.37	0.4	0.429	0.462	1	1	1	1	1	1
0.122	0.13	0.14	0.143	0.15	0.16	0.21	0.25	0.26	0.27	0.29	0.29	0.32	0.333	0.35	0.37	0.4	0.429	0.462	1	1	1	1	1	1
0.122	0.13	0.14	0.143	0.15	0.16	0.21	0.25	0.26	0.27	0.29	0.29	0.32	0.333	0.35	0.37	0.4	0.429	0.462	1	1	1	1	1	1
0.122	0.13	0.14	0.143	0.15	0.16	0.21	0.25	0.26	0.27	0.29	0.29	0.32	0.333	0.35	0.37	0.4	0.429	0.462	1	1	1	1	1	1

图 4-3 判断矩阵 F 数据结果示意

(5) 求判断矩阵 F 的最优传递矩阵 C 。

$$c_{ij} = \frac{1}{n} \sum_{k=1}^n (\lg \frac{F_{ik}}{F_{jk}}) \quad (4-10)$$

(6) 求拟优一致矩阵 D 。

$$d_{ij} = 10^{c_{ij}} \quad (4-11)$$

(7) 求解特征值与特征向量。将 D 的元素按行相乘，用方根方法求得最大特征值对应的特征向量，并进行归一化处理，进而得各因素权重。

$$u_i = \sqrt[n]{\prod_{j=1}^n d_{ij}} \quad i = 1, 2, \dots, 25 \quad (4-12)$$

$$v_i = \sqrt[n]{u_i} \quad (4-13)$$

将方根向量正规化，即得特征向量 ω 。

以上步骤应用 MATLAB 进行编程运算，计算 25 个要素权重值 w_1 如下。

$w_1 = [0.1709, 0.1493, 0.1122, 0.1034, 0.0954, 0.0765, 0.0420, 0.0279, 0.0252, 0.0228, 0.0205, 0.0205, 0.0168, 0.0153, 0.0139, 0.0127, 0.0117, 0.0109, 0.0101, 0.0070, 0.0070, 0.0070, 0.0070, 0.0070, 0.0070, 0.0070]$

由以上分析可知，改进的 AHP 方法不需要进行一致性判断，一次可得到分析结果。作为一种针对刑事案例类型的可行方法，对分析案例属性的权重有重要意义。我们将这一方法与遗传算法属性优化方法进行比较，从而确定属性优化方法的可靠性。

2. 基于遗传算法的案例属性优化分析结果

CBR 分析中，对所建立的案例库能够简捷、快速地进行案例搜索，从而完成检索到最优案例是量刑决策过程的关键环节。前述的改进的 AHP 方法可以作为一种比较方法的一种。但是，改进 AHP 方法仍然需要对权重赋值，有一定主观性。

而遗传算法在属性权重优化方面有很好的针对性。本节用遗传算法进行权重属性分析,并与改进的 AHP 方法进行比较,以确定案例属性优化的方法。

遗传算法完成属性权重分析的主要步骤:①选定二进制编码,随机产生初始种群;②确定决策变量,再应用解码方法,对经编码后的个体转换成问题空间的决策变量,并求得个体的适应值;③确定适应度函数,根据目标要求确定适应度函数;④用遗传算子对个体进行操作,按适应值的大小,从种群中选出适应值较大的一些个体进行选择、交叉、变异操作,形成新一代的种群;⑤反复执行步骤②~④,直至满足收敛判断为止。

遗传算法完成属性权重优化分析。CBR 分析中,需要对案例库中的案例进行检索,找出最相似案例,对于差别大、属性多、难以测度的司法审判案例来说,这是一个不容易解决的问题。同时,对案例特征权重向量优劣也需要进行评定,这就要解决检索精度问题。由于遗传算法在处理对象上,针对问题参数的编码集,而不是问题本身,因而可以解决许多传统方法受限制条件难以解决的问题;遗传算法在广泛的搜索空间中,可以同时在很多点进行求解,因此,不仅增加了并行处理的可能性,也减少了收敛于局部最优的可能性。而且,遗传算法基于随机规则进行搜索,使得取得最优解的可能性大幅增加,遗传算法也易于同其他技术方法有机结合,形成综合解决问题的特征。所以,在刑事案例属性优化中选用遗传算法有较好的适用性。

(1) 遗传算法适应度函数的确定。在本研究中,收集整理了 150 个刑事案例,其中检验案例库中有 45 个,记为 N_T ,检验案例库主要用于案例检索精度的分析。标准案例库中有 105 个,记为 N_R ,主要用于检索相似案例。

设检验案例 T 有 m 个属性,每一个案例的解属性有 n 个,如式 (4-14)。

为了对案例进行检索分析, 设案例的属性特征向量为:

$$T_i = \{t_1, t_2, t_3, \mathbf{L}, t_m\} \quad i = 1, 2, 3, \mathbf{L}, N_T \quad (4-14)$$

案例的属性解向量为:

$$T_{si} = \{t_{s1}, t_{s2}, t_{s3}, \mathbf{L}, t_{sn}\} \quad i = 1, 2, 3, \mathbf{L}, N_T \quad (4-15)$$

设标准案例 R 有 p 个属性, 表示为:

$$R_i = \{r_1, r_2, r_3, \mathbf{L}, r_p\} \quad i = 1, 2, 3, \mathbf{L}, N_R \quad (4-16)$$

案例的属性解向量为:

$$R_{si} = \{r_{s1}, r_{s2}, r_{s3}, \mathbf{L}, r_{sn}\} \quad i = 1, 2, 3, \mathbf{L}, N_R \quad (4-17)$$

如果 t_{sp} ($p=1, 2, 3, \mathbf{L}, n$) 表明 T 解特征向量的第 p 个属性, r_{sq} ($q=1, 2, 3, \mathbf{L}, m$) 则表明 R 解特征向量的第 q 个属性。

设 T 、 R 中解的特征向量相似程度 TRA 由 (4-18) 表述。

$$\text{TRA}_{kj} = \begin{cases} 1, & t_{sj} = r_{sj} \\ 0, & t_{sj} \neq r_{sj} \end{cases} \quad (4-18)$$

对属性求和, 可知第 k 个案例已知解与检索出的案例解的相似度:

$$\text{TRS}_k = \frac{1}{N} \sum_{j=1}^N \text{TRA}_j \quad (4-19)$$

对所有检验案例库的案例求和, 可知案例的适应度函数如式 (4-20)。

$$\text{maxFTR} = \frac{1}{N_T} \sum_{j=1}^{N_T} \text{TRS}_j \quad (4-20)$$

T 、 R 中相似度最大的案例定义为

$$\text{SIM}(T_k, R_q) = \max \text{SIM}(T_k, R_r) \quad r = 1, 2, 3, \dots, N_R \quad (4-21)$$

式中 R_q 为标准案例中相似度最大案例, 如果让解的特征确定为相等(等于 1)、不相等(等于 0)两种情况, 可将式(4-21)进一步简化。同时, 将式(4-19)、式(4-20)合并, 可得到式(4-22)。

$$\text{maxFTR} = \frac{1}{NN_T} \sum_{i=1}^{N_T} \sum_{j=1}^N \text{TRA}_j^i \quad (4-22)$$

$$\text{SIM}(T_K, R_q) = \max_{r \in R} \max \text{SIM}(T_K, R_q) = \max \frac{\sum_{i=1}^M \omega_i \cdot \text{SIM}(f_{p_i}^{T_K}, f_{p_i}^{R_q})}{\sum_{i=1}^M \omega_i} \quad (4-23)$$

式中 ω_i 为相应的属性权重。

为让相似度函数具有一般性, 将 T 、 R 中案例 T_k 、 R_k 解的相似度定义为

$$\text{FTR} = \frac{1}{N_T} \sum_{i=1}^{N_T} \text{SIM}_i(T_k, R_q) \quad (4-24)$$

(2) 基于遗传算法刑事案例属性权重优化过程。刑事案例最优权重的确定经过六个步骤, 如表 4-2 所示, 其研究框架如图 4-4 所示。

表 4-2 刑事案例各因素最优权重算法表

步 骤	主要操作过程	主要功能
步骤一	产生 m 个特征权重向量(初始权重采用改进 AHP 方法获得的权重)	组成初始群体
步骤二	给定特征权重情况下, 将测试案例输入 CBR 系统中进行测试	找出相似度最大案例 R_q
步骤三	计算 R 、 T 中 T_k 、 R_q 相应解的相似度 SIM_p	对相似度进行比较
步骤四	确定不同的权重向量对应的检索精度	精度与权重的匹配度

续表

步 骤	主要操作过程	主要功能
步骤五	从收敛条件判断其是否是最优权重向量	最优权重向量的判断
步骤六	收敛条件不满足时，对特征权重向量进行选择、交叉、变异等操作，重复以上过程	下一代特征权重向量群体的确定

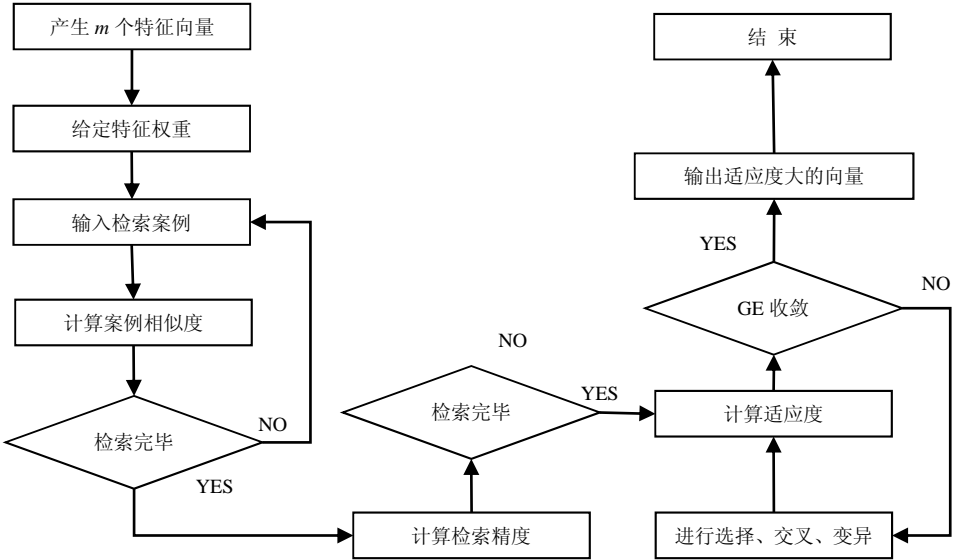


图 4-4 遗传算法流程

（3）基于遗传算法对刑事案例属性优化后的权重。运用遗传算法对案例属性进行优化，能够达到检索的准确性与可靠性。通过确定权重，按照表 4-2 的运算步骤，可得到基于遗传算法的属性权重优化值：

$$\omega_2 = [0.501, 0.501, 0.482, 0.500, 0.456, 0.496, 0.501, 0.482, 0.501, 0.407, 0.495, 0.501, 0.00, 0.314, 0.495, 0.500, 0.482, 0.501, 0.407, 0.501, 0.482, 0.0, 0.0, 0.0, 0.0]$$

3. 两种算法案例属性优化的结果比较

为了比较两种方法的结果，采用在所选定的 150 个案例集中选定 105 个案

例作为权重的优化分析数据, 45 个案例作为结果测试分析案例, 对这一方法进行验证。为了对权重优化结果进行评价, 本书采用相对均方根误差 (Relative Mean Squared Error, RMSE) 作为指标评判标准。RMSE 可定义为

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n \left(\frac{p'_i - p_i}{p_i} \right)^2} \times 100\% \quad (4-25)$$

式中, $P'_i (i=1,2,3,\dots,n)$ 为标准案例解的相对值, $P_i (i=1,2,3,\dots,n)$ 为案例模型输出解的相对值。

案例属性权重优化效果的比较主要采用三种情景检索效果比较。首先, 选择三种情景, 即无权重、改进 AHP 方法形成的权重、GA 方法形成的权重, 权重优化方法采用本节前两点所讲的方法, 权重见 ω_1 、 ω_2 ; 其次, 在专家评价方法选定的 45 个标准案例库中选择标准案例, 在测试案例库中统一用欧氏距离法选择最相似案例, 对相似案例进行解的误差分析; 再次, 分析三种情景下案例检索的误差; 最后, 完成三种情景属性权值比较分析。每种情景分析 20 个数据, 用 MATLAB 进行数据分析与作图, 如图 4-5 及表 4-3 所示。

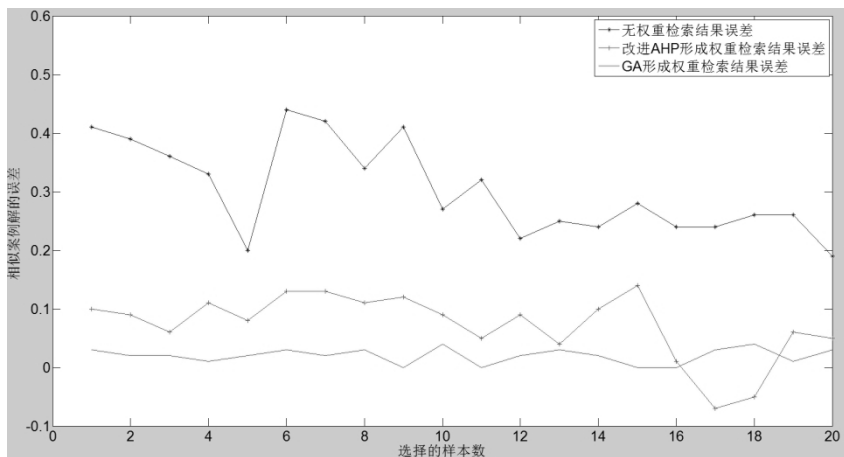


图 4-5 三种情景案例检索结果的相对误差图

表 4-3 为无权重方法、改进 AHP 方法与 GA 算法对属性权值优化的比较。

表 4-3 无权重方法、改进 AHP 方法与 GA 算法对属性权值优化比较

对比指标	无权重检索（%）	改进 AHP 方法确定权重检索（%）	GA 方法确定权重（%）
均方根误差 RMSE	12.20	9.29	5.37
最大相对误差 MRE	0.97	0.98	1.00
平均相对误差	0.86	0.69	0.39

从表 4-3 可知，GA 方法可以达到案例属性的优化，对于多案例、多属性、多要素的刑事案例库，这一方法有很好的针对性，可以提高检索过程的精度与可靠性，而且克服了其他方法存在的运算速度低、检索过程复杂、精度低等问题。另外，这一方法可以与最近邻法结合，形成精度高且方法简捷的优势。显然，这种方法组合对于指标多、内容涉及面广的刑事案例分析有重要意义。

4.2 基于 GA-KNN 算法对刑事案例检索策略研究

CBR 的关键环节在于相似案例的检索，而相似案例的检索精度取决于案例属性的权重优化值。4.1 节通过改进的 AHP 方法与 GA-KNN 综合方法对案例属性的权重进行了优化。研究结果表明，对于刑法案例，GA-KNN 方法是解决大案例、多属性、难量化刑事案例的很好的检索方法。本节将进一步论证刑事案例。

4.2.1 案例相似性的 KNN 算法

KNN 算法在案例推理中应用广泛，其主要做法是对所有案例穷尽，对案例

库中的每个案例进行比较并逐步匹配。这一算法是通过计算两案例的相似度，将相似度超出某一确定值的案例输出的过程。一般情况下，案例中的属性较多，因此，案例相似度度量确定为案例中各属性间相似度的加权和。

1. KNN 算法概述

KNN 的具体计算方法是先计算待分类样本与已知类别的训练样本间的距离或相似度，找到距离或相似度与待分类样本之间的距离或相似度与待分类数据最近的 K 个邻居；待分类样本的类别可根据这些邻居所属类别来判断，如果所属类别等同，待分类别可以划归这一类别，否则，对这些类别进行再评价，按规则确定待分类别。

如果用欧氏距离确定其相似性，设 $X、Y$ 为两个样本， $x_i(i = 1,2,3,L,n)$ 为 X 的某一分量，同理， $y_i(i = 1,2,3,L,n)$ 为 Y 的某一分量，则欧氏距离的计算公式为

$$D(x,y)=\sqrt{\sum_{k=1}^n(x_k-y_k)^2}$$

(4-26)

2. KNN 算法主要步骤

表 4-4 列出了 KNN 算法的主要步骤。

表 4-4 KNN 算法主要步骤

步 骤	主要操作过程	主要功能
步骤一	选定测试对象，确定训练集	组成初始群体
步骤二	计算每一对象与测试集中的对象的距离 公式 $D(x,y)=\sqrt{\sum_{k=1}^n(x_k-y_k)^2}$	用公式计算

续表

步 骤	主要操作过程	主要功能
步骤三	圈定距离最近的 k 个训练对象，作为测试对象的近邻	确定测试对象的近邻
步骤四	根据 k 个近邻归属的主要类别，来对测试对象分类	分类的判断

3. KNN 算法确定最大相似度案例

1) 刑事案例库及其设计

设刑事案例库中有 K 个案例，每一个案例具有 25 个属性，一般表达式为

$$R_j = R(r_{j1}, r_{j2}, r_{j3}, r_{j4}, r_{j5}, \mathbf{L}, r_{j25}) \quad (j = 1, 2, 3, \mathbf{L}, K) \quad (4-27)$$

如果给定一个待检索案例

$$X = X(x_1, x_2, x_3, x_4, x_5, \mathbf{L}, x_{25})$$

在案例库中选定一个案例

$$R_j = R(r_{j1}, r_{j2}, r_{j3}, r_{j4}, r_{j5}, \mathbf{L}, r_{j25}) \quad (j = 1, 2, 3, \mathbf{L}, K)$$

设定 K 值，一般先确定一个初始的值，然后根据试验结果反复调整到最优。

选定 $K=3$ ，按式 (4-26) 计算相应欧氏距离 $D(x,y)$ 。

选择 K 个近邻的样本，将计算出的距离降序排列，选择距离相对较小的 K 个样本作为测试样本的 K 个近邻。找出主要类别，根据 K 个近邻的类别，并应用最大概率对所查询的测试样本进行分类。所用概率指的是每一个类别出现在 K 个近邻中的比例，根据每一类别出现在 K 个近邻中的样本数量除以 K 来计算。为 K 个近邻中每一类别样本数量的集合。

2) 基于 KNN 算法的刑事案例检索结果

为了对最相似案例检索结果进行比较分析，本节采用 KNN 算法对最相似案例进行分析，确定 KNN 算法在最相似案例检索中的精度及可靠性。

根据刑事案例属性的专家解析，我们将刑事案例属性确定为 25 个指标。并给出一组待检索案例，这是一组待审判案例。指标及给出的待检索案例见表 4-5。

表 4-5 案例属性

序 号	犯罪情节	属性名称	属性具体内容	待检索案例
1	施暴方式	M1	打	0
		M2	刺	1
		M3	其他	0
		M4	人数	1
2	使用工具	M5	拳脚	0
		M6	钝器	1
		M7	锐器	0
3	袭击部位	M8	胸部	0
		M9	腹部	1
		E10	头部	0
		E11	四肢	0
		E12	其他	0
4	加害结果	E13	D4	0
		E14	D6	0
		E15	D8	1
		E16	D11	0
5	从轻环节	M17	自首	0
		M18	积极救治	0
		M19	认罪	0
		M20	悔罪	1
		M21	积极赔偿	0

续表

序 号	犯罪情节	属性名称	属性具体内容	待检索案例
6	刑罚类型	M22	刑罚 E1	1
		M23	刑罚 E4	0
		M24	刑罚 E5	1
		M25	刑罚 E6	0

对表 4-5 给出的待检索案例特征，用 KNN 算法进行计算，选出的相似案例见表 4-6，其平均量刑为有期徒刑 4 年。

表 4-6 KNN 方法检索得到的相似案例情况

检索组	案 例 数	KNN 方法相似度	平均相对误差（%）
第一组	26	0.78	8.23
第二组	40	0.67	7.37

4.2.2 案例相似性的 GA-KNN 算法

为了提高检索的精度与可靠性，针对刑事案件案例多属性、多案例、难量化的特点，本节探讨 GA-KNN 相结合的方法，通过与 KNN 方法的对比，进一步提高这种方法对案例检索的可靠性。

1. 案例相似性 GA-KNN 方法

KNN 算法的缺陷在于运算量大，且得到多个案例，对于多案例难度增大，精度也难以达到要求。刑事案例应用 GA-KNN 算法，可以很好地解决多案例问题，通过 GA 的收敛搜索方法，确定相应条件，可以达到很快找到最优案例的目的，为决策提供很好的支持。按照 4.1.3 小节第二点的论述，可以通过表 4-7 的步骤进行相似案例的检索。

表 4-7 刑事案例 GA-KNN 检索步骤表

步 骤	主要操作过程	主要功能
步骤一	产生 m 个特征权重向量(初始权重采用改进 AHP 方法获得的权重)	组成初始群体
步骤二	给定特征权重情况下, 将测试案例输入 CBR 系统中进行测试	找出相似度最大案例 $SIM(T_k, R_q)=\max SIM(T_k, R_r)$
步骤三	计算 $R、T$ 中 $T_k、R_q$ 相应解的相似度 SIM_p	对相似度进行比较 $SIM_p(T_k, R_q) = \frac{1}{N} \sum_{j=1}^N SIM_p(r_{sj}, t_{sj})$
步骤四	确定不同的权重向量对应的检索精度	精度与权重的匹配度 $TRS_k = \frac{1}{N} \sum_{j=1}^N TRA_j = \frac{1}{N} \sum_{j=1}^N SIM_p(r_{sj}, t_{sj})$
步骤五	从收敛条件判断其是否是最优权重向量	最优权重向量的判断
步骤六	收敛条件不满足时, 对特征权重向量进行选择、交叉、变异等操作, 重复以上过程	下一代特征权重向量群体的确定 GA 运算过程

2. 案例相似性 GA-KNN 方法运算结果

输入表 4-5 的待检索案例, 可以得到最相似案例集, 如表 4-8 所示。同样量刑为有期徒刑 4 年, 但得到的案例相似程度和检索精度明显高于 KNN 方法。

表 4-8 GA-KNN 方法检索得到的相似案例情况

检 索 组	案 例 数	GA-KNN 方法相似度	平均相对误差 (%)
第一组	129	0.98	1.13
第二组	44	0.92	2.32

4.2.3 案例相似性的 KNN 方法与 GA-KNN 算法的比较

为了提高案例检索的精度与可靠性, 采用 GA 与 KNN 方法相结合进行案例的检索。本节对两种检索方法进行比较, 从而分析 GA-KNN 组合方法在刑事

案例检索中的精度与可靠性。首先，选择两种方法对 105 个测试案例进行有权重的相似案例检索，待检索案例属性见表 4-5；其次，在专家评价方法选定的 45 个标准案例库中选择标准案例，在测试案例库中用 KNN 方法与 GA-KNN 方法进行最相似案例操作，对相似案例进行解的误差分析；再次，分析两种情形下案例检索的误差；最后，完成两种情形案例检索误差的比较分析。每种情形分析 20 个数据，用 MATLAB 进行数据分析与作图，如图 4-6 及表 4-9 所示。

GA-KNN 算法明显比 KNN 算法的精度与可靠性得到了大的提高，表 4-9 是两者结果的比较。GA-KNN 算法权重优化值为

$$\omega_3 = [0.010, 0.015, 0.011, 0.016, 0.010, 0.020, 0.022, 0.022, 0.020, 0.012, 0.025, 0.013, 0.04, 0.095, 0.149, 0.170, 0.007, 0.007, 0.007, 0.007, 0.007, 0.007, 0.027, 0.076, 0.130]$$

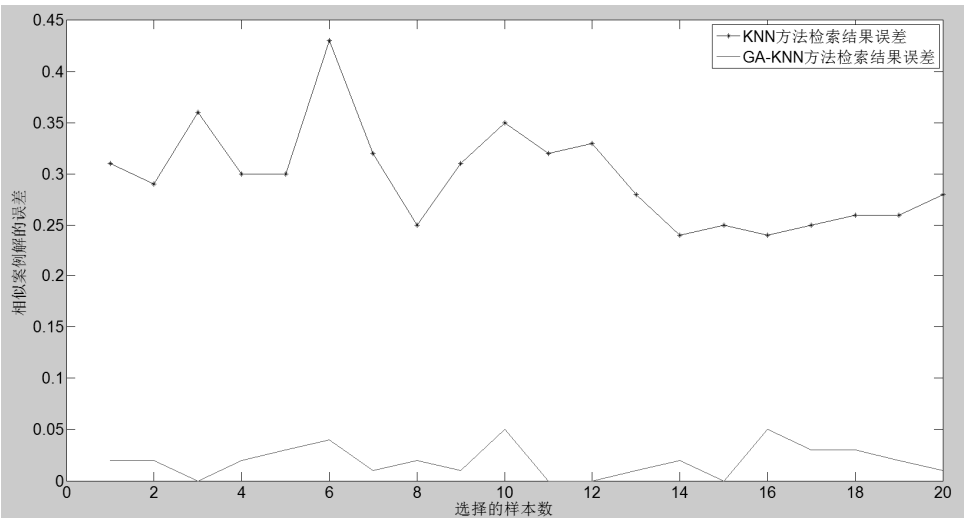


图 4-6 KNN 与 GA-KNN 案例检索结果的相对误差图

表 4-9 KNN 方法与 GA-KNN 算法对检索结果的比较

	KNN 方法	GA-KNN 方法
均方根误差 RMSE	0.30794	0.025
最大相对误差 MRE	0.96875	0.324
平均相对误差	0.87186	0.833

从表 4-9 可知，GA-KNN 方法可以使案例检索精度大幅度提高，同时与最近邻法结合，达到精确度提高，均方误差、最大相对误差降低的目的。因此，GA-KNN 方法是刑事案例决策分析的重要依据。在第 5 章的推理集成决策机制研究中，就是以 GA-KNN 方法检索出来的，针对特定决策需要的最优案例库为基础，展开分阶段进行量刑决策研究的。

在案例属性优化与案例检索过程中调用 MATLAB 算法，基于 SQL 数据库，完成了检索系统，实现了案例属性优化与案例检索，案例检索对比分析的可视化与多次数据分析。所实现的数据分析界面如图 4-7 所示。



图 4-7 案例检索系统界面图

4.3 本章小结

本章通过分析刑事案件属性的特征及属性优化的必要性,针对刑事案例属性优化问题进行了论证,建立了基于 GA-KNN 方法对刑事案例特征属性优化方法。用改进的 AHP 方法进行了比较,拓宽了属性优化方法,构建了刑事案例中相似案例的确定方法,获得了决策推理的最优案例集。针对司法案例的多案例、多属性、多取值案例特征,首次实现了同类案例相似性检索的精度和可靠性,为案例的决策推理提供了可靠的保证。这一研究对以往司法类案例检索方法与路径研究有所突破,实现了案例检索的针对性、完整性、快捷性、应用性。

刑事案件推理集成决策模型研究

法官量刑需要针对特定刑罚类型的符合人类思维习惯的推理方法，推理的核心算法需要经过相似案例库的训练，确保推理模型的科学、可靠。本章研究了基于刑事案件本体的规则推理和案例推理方案，从基于本体的案例特征提取、推理模型的训练，使用三个阶段构建了方案框架，具体的推理过程是基于特征选择、决策树模型、神经网络模型等数据挖掘算法，构建支持量刑决策的推理集成模型^[123]实现的。经过实验验证，方案推理结果准确度较高，能够为审判人员提供一定的数据支撑，具有较高的可行性。

5.1 基于本体的法律推理体系

在法律推理研究领域，得出判决结果的逻辑推理过程一直是研究的重点和难点。从推理过程所依靠的“事实”的角度分类，可以分为基于规则的推理和

基于案例的推理。在传统的审判工作中，大多采用规则推理。随着 RBR 和 CBR 模型及其相关人工智能技术的日益发展，加之以对两种推理模型各自优缺点的认识，将这两种推理模型相结合成为一种重要的趋势。Jim Prentzas 总结了一些将 RBR 和 CBR 相整合的推理系统原型，包括独立式和耦合式两种，在法律领域也有应用尝试。法律推理本身是一种缜密的逻辑思维加上丰富的经验的过程，这里的“缜密的逻辑”就是法律法规，“丰富的经验”就是蕴含在判决书中的法官量刑的智慧。具体到中国法律体系，法律法规中所蕴含规则的“条件”和“结论”并不十分清晰，大都需要司法实践中遇到问题，就问题给出的司法解释指导一类问题，但司法解释不但不能穷尽所有法律，而且也不能就一类问题涵盖得那么具体。而案例是审判历史过程中形成的、已经为人所接受的、具有法律效力的裁判。从案例中提取案例特征进行基于案例的推理，则需要对决策因素进行合理的组织，以确保因素之间语义关系的准确性，这一点在前面通过本体对刑事案件内容进行了规范表达。

在实际应用中，传统的研究方法存在以下两个短板，一是建立基于案例特征的规则集需要较大工作量，案例特征的数量（维度）难以估量，特征与判决结果的对应关系需要专业人士的权衡；二是目标案例与历史案例相似关系的比较可能无法完全用线性关系表征，且若涉及案例中的所有因素，这一比较过程也会相当耗时。

为此，本书所用的方法是利用第 4 章以刑罚类型为单位建立的案例库，组合应用特征选择、决策树及神经网络等技术，挖掘相应刑罚类型的法律推理规则，再嵌入案例库中，用于向法官推荐高度相关的参考基础案例和推荐量刑；法官依据参考案例和推荐量刑做出目标案例量刑；当量刑结果在刑罚类型的量刑区间内而又与推荐案例量刑结果不完全相同时，则此目标案例自动进入案例

库，案例库则在一定周期内，对决策树模型和神经网络模型进行自动训练，实现案例库和决策模型的动态更新。

5.2 刑事案件推理决策模型中的相关算法

在研究进行法律推理规则挖掘中，使用特征选择、决策树、神经网络等数据挖掘算法，以下内容对这些技术基础进行介绍。

5.2.1 特征选择

特征选择，指从高维的特征集合中，依据某种标准进行评估，从而选择输出性能最优的特征子集。其目的就是找到保持数据集较为主要特征的低维数据集，通过低维数据的分析获得相关高维数据特性，以达到简化分析、获取数据有效特征及可视化数据的目标。

特征选择的一般流程可用图 5-1 表示。首先，从特征全集中根据需要产生特征子集，然后使用评价函数对这一特征子集进行评价，用停止准则与评价结果进行比较。如果评价结果比停止准则好则停止选择，否则产生另一组特征子集，继续进行特征选择过程，直到不满足停止准则为止。特征子集还要验证其有效性。

综上所述，特征选择的一般过程包括：产生过程、评价函数、停止准则、验证过程。

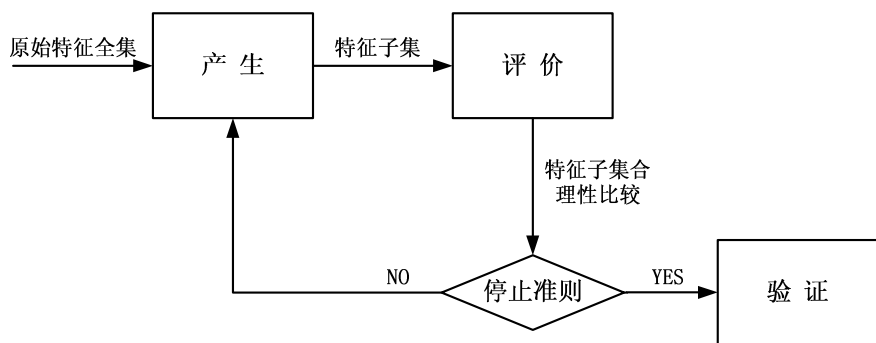


图 5-1 特征选择的一般流程

(1) 产生过程。案例特征产生过程主要进行特征子集的搜索，负责为评价函数提供特征子集。搜索特征子集的过程有多种，算法主要分为完全搜索、启发式搜索及随机搜索 3 类。

(2) 评价函数。评价函数是评价一个特征子集好坏程度的一个准则。根据其工作原理，主要分为筛选器和封装器两大类。一般在预处理过程中，筛选器通过分析特征子集内部特点衡量特征子集的优劣；封装器实质上是分类器，使用选取的特征子集对样本集分类，衡量特征子集优劣的基本标准是分类精度。

(3) 停止准则。停止准则与评价函数相关，即为阈值，当评价函数的值达到该阈值后搜索就停止。

(4) 验证过程。验证过程即验证选择的特征子集的有效性。

如上所述，特征选择的主要目的是寻找对输入变量预测有积极贡献的重要变量。变量的重要性可以从两个方面联合考察：第一，从变量本身考察；第二，从输入变量与输出变量的相关性角度考察。

1. 从变量自身考察特征

从变量本身看,重要变量应是携带信息较多,即变量值差异较大的变量。从统计学的角度看,测度数值型变量取值离散性的指标是标准差或变异系数。标准差越大,说明变量取值的离散程度越大,反之越小;变异系数在消除数量级影响的情况下,便于对多个变量的离散程度进行对比。通常的考察标准有:

(1) 如果某数值型变量的变异系数小于某个标准值,则该变量应视为不重要变量。

(2) 如果某数值型变量的标准差小于某个标准值,则该变量应视为不重要变量。

另外,对分类型变量取值差异性的测度的指标和标准是:

(1) 对某分类型变量,计算各个类别值的取值比例。如果其中的最大值大于某个标准值,则该变量应视为不重要变量。

(2) 对某分类型变量,计算其类别值个数。如果类别值个数与样本量的比大于某个标准值,则该变量应视为不重要变量。

2. 从输入变量和输出变量相关性的角度考察特征

特征选择的过程,即为对预测变量对于目标变量重要程度的计算并依据重要程度的大小对预测变量进行排序和筛选。使用 importance 指数标识该重要程度。这里,使用假设检验的相关原理做预测变量与目标变量的相关分析,提出原假设为:

H_0 : 变量 X 与变量 Y 不相关

在样本数据集中, 计算某一事件在 H_0 假设下发生的概率 (即 p 值), 若 p 值较小, 则有一定理由拒绝 H_0 , 即变量 X 与变量 Y 具有一定的相关性。

因此, importance 指数可以用如下的方式计算:

$$\text{importance} = 1 - p$$

根据预测变量与目标变量的数据特征, 即离散型或连续型, 计算 importance 指数的方法也有所区别, 具体到刑事案件案例推理的实际特征, 这里主要介绍预测变量与目标变量为离散型的计算方法。

使用基于皮尔逊 χ^2 的方法度量变量之间的相关性, 其中的符号定义如下:

X ——离散预测变量, 有 I 个不同的取值;

Y ——离散目标变量, 有 J 个不同的取值;

N ——总样本数量;

N_{ij} —— $X = i$ 且 $Y = j$ 的样本数量;

$N_{i \cdot}$ —— $X = i$ 的样本数量;

$N_{\cdot j}$ —— $Y = j$ 的样本数量。

令卡方统计量为:

$$\chi^2 = \sum_{i=1}^I \sum_{j=1}^J \frac{(N_{ij} - \hat{N}_{ij})^2}{\hat{N}_{ij}} \quad (5-1)$$

其中, \hat{N}_{ij} 为频数期望值, $\hat{N}_{ij} = \frac{N_{i \cdot} \cdot N_{\cdot j}}{N}$ 。 χ^2 值越大, 表示预测变量与目标

变量之间的关联程度越高。

预测变量 X 的重要程度定义为:

$$\text{importance} = 1 - \text{Prob}(c_a^2 > c^2) = 1 - a \quad (5-2)$$

在计算 χ^2 之后, 依据 χ^2 的值查询 χ^2 分布表, 从而得到显著性水平, 即 p 值, 即可求出 **importance** 指数。最后, 根据以下的规则, 对所有的预测变量排序:

- (1) 根据 p 值的大小对预测变量进行升序排列。
- (2) 若存在相同 p 值, 依据 χ^2 统计量的大小降序排列。
- (3) 若仍存在相同的 χ^2 统计量, 按自由度 d 的大小升序排列。
- (4) 若仍存在相同的 d 值, 则按照数据的自然顺序排列。

对于连续型的预测变量, 使用 F 统计量的 p 值计算 **importance** 指数:

$$\text{importance} = 1 - \text{Prob}(F(J-1, N-J) > F) = 1 - a \quad (5-3)$$

F 统计量的计算公式为:

$$F = \frac{\sum_{j=1}^J N_j (\bar{x}_j - \bar{x})^2 / (J-1)}{s_j^2 (N_j - 1) / (N - J)} \quad (5-4)$$

服从分子自由度 $K_1 = J - 1$, 分母自由度 $K_2 = N - J$ 的 F 分布。其中, \bar{x}_j 为对所有 $Y = j$ 的样本, 预测变量 X 的样本均值; \bar{x} 为预测变量 X 的总体均值。

5.2.2 决策树

决策树是一种倒立的树结构, 它由内部节点、叶子节点和边组成, 如图 5-2

所示，其中，最上面的节点为根节点。决策树的构造过程需要使用训练集，每个例子使用若干属性（或特征）和类别标记来描述。构造决策树的目的是寻找属性和类别之间的关系，从而使用“关系”预测未知记录的类别。将这一具有预测功能的系统称为决策树分类器。

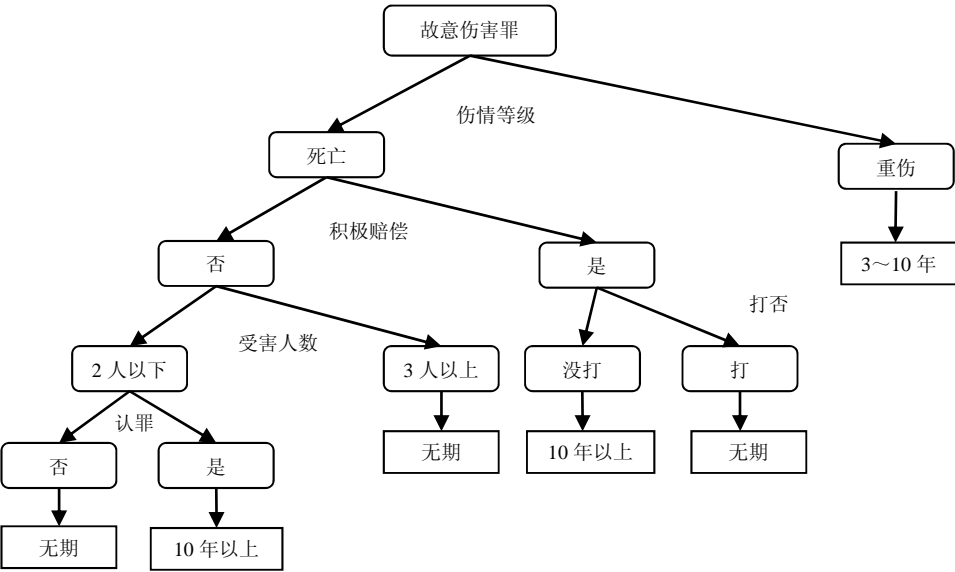


图 5-2 决策树示例

构造一个决策树分类器通常分为两步，即树的生成和剪枝。其中，树的生成是采用自上而下的递归方法。以多叉树为例，它的构造思路为：若训练例集合中的所有例子是同类的，则称之为叶子节点，节点内容即为该类别标记；否则，根据策略选择某属性，按照属性的若干取值，将训练集划分为多个子集，使每个子集中的例子都在该属性上具有同样的属性值。然后，再依次递归处理各个子集。这种思路实际上就是递归分类。由此看出，在节点处所选的属性不同，则会构成不同的决策树，而不同的树的预测精度一般也是不同的。因此，在决策树的构成步骤中，核心是找出节点的属性和如何对属性值的划分。图 5-3 示意决策树的生成过程。

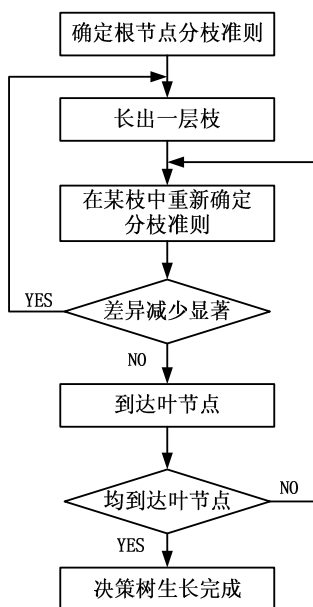


图 5-3 决策树生成过程

其中，差异减少是否显著是指分组样本中输出变量取值的差异性是否随决策树的生长，即分组的进行而显著减少。有效的决策树分枝（分组）应当使枝（组）中样本的输出变量取值差异迅速减少，尽快趋同。

若训练集的数据中存在有噪声信息，如某些记录的属性值错误、类别标记有误等，则这种完全符合就造成了过分拟合，反而降低了模型的预测性，因此需要进行决策树模型的第二步——剪枝。剪枝即为剪去那些不会增加决策树错误预测率的分枝。

决策树可以定义为一种推理过程，它使用类似于树形的数据结构来描述数据集，适用于解决分类问题。常见的决策树算法有 C5.0、CART、CHAID 等。

本书中，使用 CART 决策树算法，即分类和回归树（Classification and Regression Trees, CART）。在 CART 算法中，当类标号属性是离散型数据时，

生成的结果即为分类树。对于给定的样本集 S ，使用该算法建立决策树可以划分为如下三个步骤：一是生成最大树；二是树的修剪；三是子树评估。

1. 生成最大树

决策树分裂的原则为，当原始数据集分裂为很多子集后，子集中的数据需要尽力属于同一类别。使用杂度（Impurity）度量数据集随机性的程度，并采用 gini 指标衡量。设决策树某一节点 t ，这一节点的数据集是 S ，该数据集包含 s 个样本，它的类标号属性包含 m 个不同的取值，即 m 个相异的类 $C_i (i = 1, 2, \dots, m)$ 。设属于类 C_i 的样本有 s_i 个。则该节点的 gini 指标计算公式如下：

$$\text{gini}(t) = 1 - \sum_{i=1}^m (p(C_i | t))^2 \quad (5-5)$$

其中， $p(C_i | t) = \frac{s_i}{s}$ ，表示 C_i 类在数据集 S 中的概率。因为 CART 算法所生成的是一棵二叉树，节点 t 分裂后会产生子节点 t_L 和 t_R ，则在这一过程中的杂度削减为

$$j(S, t) = \text{gini}(t) - p_L \text{gini}(t_L) - p_R \text{gini}(t_R) \quad (5-6)$$

其中， p_L 是 t 中的样本被划分至 t_L 中的概率， p_R 是 t 中的样本被划分至 t_R 中的概率。选择杂度削减量最大的属性作为分类属性，即确定最佳分支方案。

上述方法用于对离散型属性进行划分，对于连续型的类标号属性，采用最小平方误差（Least-Squared Deviation）。

2. 树的修剪

经过上述步骤，能够得到最大树 T_{\max} ，需要进一步地对其中多包含的众多

叶子节点进行“修剪”，降低决策树的复杂度。使用代价复杂度这一指标，综合衡量误分类损失和复杂度：

$$R_a(T) = R(T) + a \left| \mathcal{L}(T) \right| \quad (5-7)$$

其中， $R_a(T)$ 为决策树 T 的复杂度； $R(T)$ 是其误分类损失； $|\mathcal{L}(T)|$ 为叶子节点数量； a 为复杂度系数，即每增加一个节点的代价复杂度。随着 a 不断增大，节点 T 是首个使得 $R_a(\{t\}) = R_a(T)$ 成立的节点，称其与决策树具有最弱连接，这类节点即为修剪的对象。重复执行上述过程，直到只有一个根节点，修剪的结果为一个数的序列 T_1, T_2, \dots, T_k ，它们均为 T_{\max} 的子树，且节点数递减。

3. 子树评估

通过子树评估寻找上述子树序列中的最优子树。评估方法是计算各棵子树的误分类损失，计算公式如下：

$$R(T) = \frac{1}{N} \sum_{i \neq j} c(i|j) N_{ij} \quad (5-8)$$

其中， N 是数据集中的样本个数， $c(i|j)$ 是把 j 类误分为 i 类的损失， N_{ij} 表示将 j 类误分为 i 类的样本数量。基于 1SE 规则 (One Standard Error)，可以计算其标准误 $SE(R(T))$ ，将误分类损失落在 $[\min_k R(T_k), \min_k R(T_k) + SE(R(T))]$ 区间中的树挑出，并选出其中节点数量最少的树即为最终的决策树模型。

在 CART 中，子树评估的筛选公式为：

$$R(T_{\text{opt}}) \leq \min_k R(T_k) + m \times SE(R(T)) \quad (5-9)$$

满足上述公式的 T_{opt} 即为最优树。其中， m 默认为 1。

5.2.3 神经网络

人工神经网络（Artificial Neural Network, ANN），简称神经网络（Neural Network, NN），它是一种模仿生物神经网络的结构和功能的数学模型，由大量的人工神经元连接进行计算，是一种自适应系统，主要是解决非线性数据关系建模问题，常用来对输入和输出间复杂的关系进行建模，以探索数据之间的内在规律。神经网络的每个节点代表一种特定的输出函数，称为激励函数。每两个节点间的连接代表一个对于通过该连接信号的加权值，称为权重，这相当于人工神经网络的记忆。网络的输出则因不同神经网络的连接方式、权重值和激励函数的不同而不同。

多层感知器是目前应用得较为广泛的前向神经网络之一，由输入层、隐含层、输出层组成，如图 5-4 所示。其中，隐含层可以为一层，也可为多层。它所使用的激活函数是 sigmoid 函数：

$$S(x) = \frac{1}{1 + e^{-x}} \quad (5-10)$$

其中， x 为神经元所有输入的加权和， $S(x)$ 为其输出。

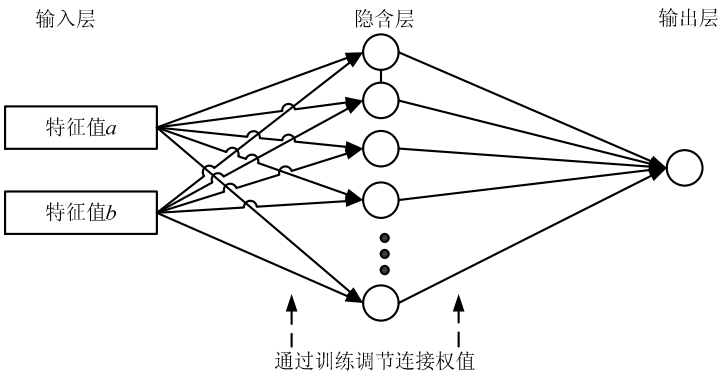


图 5-4 神经网络示例

多层感知器每一神经元均需完成以下两种计算：一是计算各神经元输出的函数信号，其表征为全部输入信号和权值的连续非线性函数；二是计算每一神经元反向传播的误差信号，以及依据误差信号计算的权值。

神经网络训练开始时，全部权值被随机地设为 $[-0.5, 0.5]$ 区间中的某值。将记录中每个属性的值作为神经网络的输入，各神经元依据输入和权值计算其输出值，输出值的计算公式为：

$$a_i = s(\sum_j w_{ij} o_j) \quad (5-11)$$

其中， w_{ij} 为神经元 j 到 i 的权值， o_j 是神经元 j 的输入，使用激活函数计算神经元 i 的输出 a_i 。逐层计算各神经元的输出后，即得到输出层的预测值，将其与实际值比较得到预测误差，继而对各权值进行更新。

首先要计算每个神经元的反向传播误差。

(1) 若神经元 j 属于输出层，则其反向传播误差为：

$$d_{pj} = (t_{pj} - o_{pj})s'(a_j) = (t_{pj} - o_{pj})o_{pj}(1 - o_{pj}) \quad (5-12)$$

其中， t_{pj} 为记录 p 的期望输出，即实际值； o_{pj} 为神经元的预测输出，即计算得到的输出值； $s'(a_j)$ 是激活函数的一阶导数在 $x=a$ 处的值。

(2) 若神经元 j 属于隐含层，则其反向传播误差为：

$$s_{pj} = s'(a_j) \sum_k d_{pk} w_{jk} = o_{pj}(1 - o_{pj}) \sum_k d_{pk} w_{jk} \quad (5-13)$$

其中， k 表示 k 个反向输入的误差； w_{jk} 为神经元 j 与 k 相连的权值； o_{pj} 为神经元的预测输出，即计算得到的输出值。

在计算完反向传播误差后,计算突触权值的改变量 Δw 。神经元 i 到 j 的权值改变量为:

$$\Delta w_{ij}(n+1) = h d_{pj} o_{pi} + a \Delta w_{ij}(n) \quad (5-14)$$

其中, h 为学习率参数, $0 < h < 1$; d_{pj} 为神经元 j 的反向传播误差; o_{pi} 为神经元 i 的输出; a 是动量常数,通常为正数; $\Delta w_{ij}(n)$ 为 w_{ij} 在上一次的改变量。

之后,每个权重值被更新为:

$$w_{ij}(n+1) = w_{ij}(n) + \Delta w_{ij}(n+1) \quad (5-15)$$

当满足预先设定的停止准则后,形成最终的神经网络模型。

5.3 刑事案件推理集成决策模型研究

5.3.1 推理集成决策模型建立意义

从审判量刑决策主体的角度来看,使用案例库,实际上是一个针对多变量决策问题进行动态选择和判断的过程。作为法官,首先要判断案件的罪名类别,再根据案件事实判断属于某一罪名下的哪条哪款,即刑罚类型。根据刑罚类型相同的以往的相似案例进行提出量刑,但即使是相似案例,它们之间、它们与目标案件之间的案情都存在区别,其中既要根据法规,也要依赖案例,依法官思维,提供决策支持必须在不同过程提供推理支持。

将传统的一步到位的量刑方法转变成“刑”和“期”两步完成。通过“刑”,也就是具体刑罚类型的判断,可以据此在全集案例库中选择判决结果在同一个

刑罚区间里的案例，作为讨论刑期的具体案例，这样可以大大缩小相似案例的数量，使基于这些相似案例进行刑期推理更具有可信度。量刑过程中，每一个特征的变化都会影响法官的裁量，量刑决策过程实际上是一个非线性的动态决策过程，为此，用相同刑罚区间的相似案例训练出对应的神经网络模型，然后以目标案件的案情特征为输入，进行“期”的推理。

整个过程，先是规则推理，后是案例推理，二者相互级联。级联体现在：通过决策树模型对同一罪名下的案例进行分类，实现细化后的刑罚类型与案例特征规则集进行对应的关系，形成同一个刑罚类型对应一个或多个案例特征规则集。这样当一个新的案件来时，可以直接对应案例特征规则集，如果属于这个集合中的一个，那么量刑就建议为对应的刑罚类刑。刑罚类型确定后，可以作为刑期推理选择案例的条件，以此选择对应的神经网络模型，对刑期进行推理。参考推荐结果，法官做出最终量刑，当量刑结果与参照案例量刑结果不同时，此目标案例进入案例库，实现案例库自动更新，对应的决策树模型和神经网络模型在一定周期内进行自动训练，实现案例库和决策模型的动态更新。整个推理过程和知识表示如下。

目标案例 i 的特征组合： a_i

刑罚类型 E_j 的案例库特征规则集： R_j ，它的元素是 R_{jk} ， $k=1,2,\dots,N$ 。表示刑罚类型 E_j 有 N 个案例特征组合，即有 N 种犯罪特征组合的量刑区间相同。

刑罚类型 E_j 对应的刑期推理的神经网络模型 N_j ，则推理知识逻辑关系为：

IF a_i 属于 R_j

THEN 目标案例 i 的刑罚类型为 E_j

以 a_i 为 N_j 输入条件推理刑期, 并推荐与该刑期相同且刑罚类型相同的参考案例。

IF 法官对目标案例 i 量刑与参考案例相同

THEN 目标案例 i 不进入刑罚类型 E_j 案例库

ELSE 目标案例 i 进入刑罚类型 E_j 案例库

ENDIF

ELSE 目标案例是新类型案件, 法官自由裁量后进入对应的刑罚类型案例库

ENDIF

本章后续将上述推理知识表示模型化, 并将模型重点放在刑罚类型和刑期的推理和二者集成上面, 对于案例库更新在模型使用阶段自动实现。当然, 在推理过程中, 涉及从案例库中选择关键特征问题, 这里采用与判决结果相关性作为条件进行判断。

5.3.2 推理集成决策模型设计

刑事案件推理集成决策模型主要应用了前面讨论的特征选择、决策树和神经网络。

决策树模型, 主要用于分类预测, 就是通过向可靠的历史数据学习, 使模型具备对未来新数据或新案例进行分类预测的能力。模型的这种能力源于对现有数据所包含的分类规律的归纳和提炼, 同时, 只有通过有指导的学习才可能实现可靠决策树模型的建立。这里“有学习的指导”是指数据除包含作为输入

角色的变量外，一定还包含作为输出角色的输出变量，且输出变量的取值在现有数据上是已知的。比如，本书研究的量刑决策问题使用的按刑罚类型分类的相似案例库，其中的案例不仅包括量刑决策的输入条件，如案情，同时也包括案例的判决结果。它指导着模型的学习，使模型能够理解怎样的输入变量取值或组合，能得到怎样的分类结果。具体本书采用的 CART 决策树模型，就是通过现有案例的学习，使其具备对目标案件确定量刑类别的预测能力。这里“案例”包括案情、判决依据和判决类型，通过这种有指导的学习，最终得到针对特定罪名的决策树分类模型，使其支持案件决策过程。

如图 5-5 所示是本研究采用的研究思路，刑事案件本体及其案例库的建立，如前文所述，以上述研究为基础，选择刑事案件本体及其案例库作为组织案例特征，从中抽取已经做了标注的刑事案件案例集合，以此作为基于本体的刑事案件案例推理和相似案例的推荐研究的数据基础。

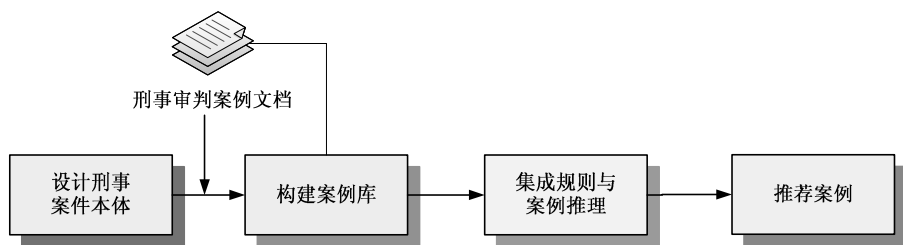


图 5-5 推理集成方案路线图

进行案件推理研究时，主要使用刑事案件本体中的案情类。如图 5-6 所示，为故意伤害罪的案情类的组织结构和语义关系。

案情类包括起因、经过、结果、案情相关类 4 个子类，前 3 个子类用于存储刑事案件的起因、经过、结果等相关信息；案情相关类主要用于对经过子类 and 结果子类进行描述，由于这些用于表述的子类与经过子类和结果子类之间并

没有父类-子类的关系，所以设计案情相关类对它们进行组织，如故意伤害罪的案情相关类包含袭击部位、作案工具、伤情等级、从轻情节、从重情节 5 个子类。

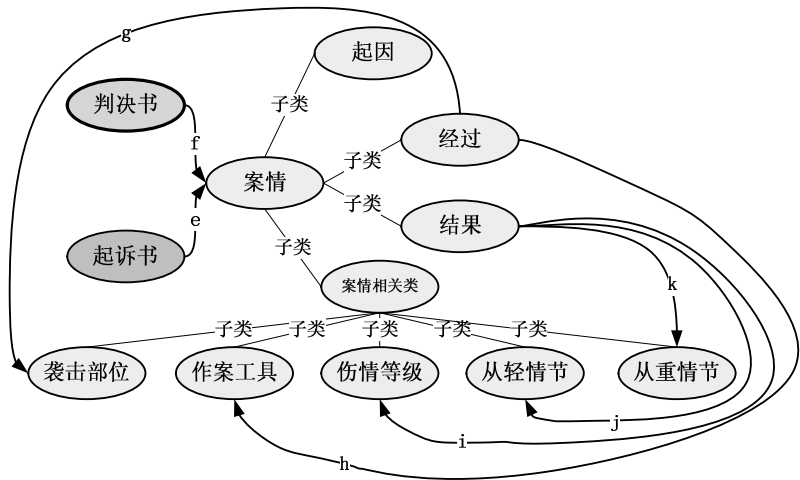


图 5-6 案情类结构

语义关系方面，图 5-6 中的 e 箭头表示公诉机关在起诉书中所指控的案情；f 箭头表示合议庭在对案件的审理过程中所查明的案情，上述两个语义关系对应案情类的不同实例。g、h 箭头表示案件经过中的袭击部位和所使用的作案工具；i、j、k 箭头表示故意伤害案件所造成的伤情等级（包括轻伤、重伤、死亡等）、从轻情节（如救治、自首、赔偿等）和从重情节（如逃逸等）。

刑事案件审理的推理模型，就是已判决刑事案件案例集合中案情及语义特征与判决结果之间的关联规则模型，并使用该模型，对未判决的刑事案件（目标案例）推荐参考案例和判决结果，为法官判案提供基于案例的决策支持。

经过比较研究，本书采用如图 5-7 所示的决策推理挖掘模型。挖掘算法的过程主要采用“级联”的思想，首先通过特征选择算法选择影响判决类型的关

键特征；其次，基于目标案例的类型（罪名）判决关键特征，使用决策树模型推断目标案例的判决类型，即刑罚类型；然后，针对每一类判决类型，通过特征选择算法选择影响刑期判决的关键特征，基于目标案例的刑期判决关键特征，使用神经网络模型推断目标案例的判决刑期；最后，基于推理的案例类型及其对应的刑期推荐参考案例。

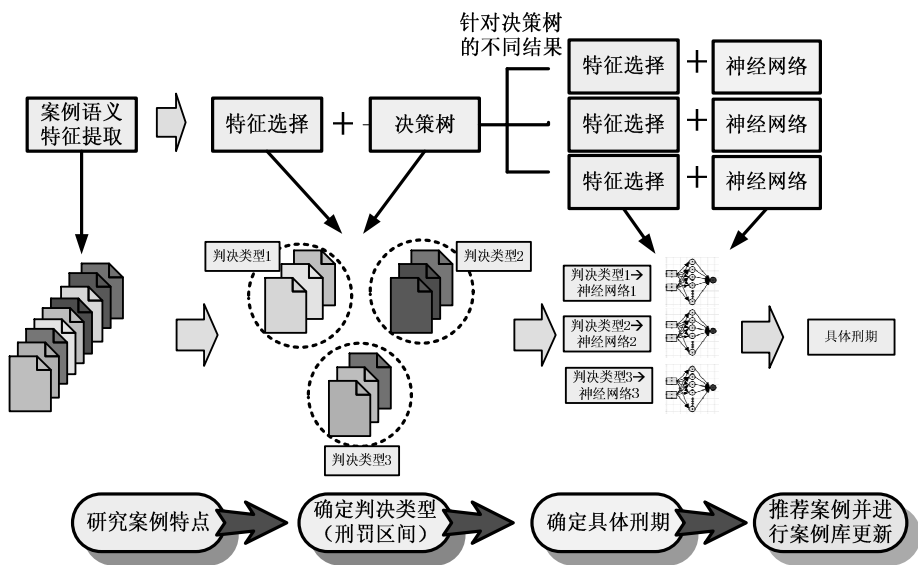


图 5-7 推理算法框架

推理挖掘算法框架的伪代码如下所示，分为两个阶段，一个是推理模型训练学习阶段，一个是推理模型使用阶段。

在推理模型训练阶段，从刑事本体案例库中抽取刑事案例存放于集合中，这些案例已在案例库中对案情及其语义特征、案例判决类型和刑期进行了标注。

训练学习过程分为两部分，第一部分是案件判决刑罚类型推理决策模型训练，在上述案例集合上，使用特征选择算法确定影响案例类型判决的特征集合，然后使用决策树算法进行训练，确定案例判决类型的决策树推理模型。第二部

分是案件判决刑期推理决策模型训练，先将上述案例集合按判决类型划分为相应的子类型集合，针对每一个子类型集合，使用特征选择算法确定相应的影响案例刑期判决的特征集合，然后使用神经网络算法进行训练，确定相应的案例刑期判决神经网络推理模型。

在推理模型使用阶段，首先，在案件判决刑罚类型推理决策模型使用阶段，先从目标案例中抽取与判决刑罚类型相关的案件特征，利用刑罚类型推理决策模型，推断出目标案例的判决刑罚类型。

然后，依据推断的判决刑罚类型，选择该刑罚类型对应的神经网络模型，同时，从目标案例中抽取相应的与判决刑期相关的案件特征作为神经网络的输入，来推断目标案例的判决刑期。

伪代码：

```
//模型的训练阶段
//案例类型判决推理模型训练数据准备
KL=刑事本体案例库案例集合；
//案例类型判决推理模型训练
KF = FD(KL)； //使用特征选择算法 FD 确定案例判决类型的关键特征集合
KFM = DT(KF, KL)； //使用决策树算法确定案例判决类型推理模型
//案例刑期判决推理模型训练数据准备
划分 KL 为  $KL_1, KL_2, \dots, KL_n$  ； //对 KL 集合按判决类型划分为 n 个子类集合
//案例刑期判决推理模型训练
for(j = 1; j <= n; j++)
{
    KKFj=FD(KLj)； //再次使用特征选择算法 FD 确定 KLj 的关键特征集合为 KKFj
    //训练神经网络模型 NNMj 为 KLj 的关键特征 KKFj 与刑期 TERMx 的推理模型
    NNMj = NN(KKFj, TERMx)；
}
```

```

// 模型的使用阶段
    // 目标案例判决类型推断
    根据模型训练阶段确定的关键特征，从目标案例提取关键特征  $KF_{TC}$ ；
    // 确定目标案例的判决类型
     $KFM_{TC} = KFM(KF_{TC})$ ;
    // 目标案例刑期类型推断
    switch( $KFM_{TC}$ )
    { // 针对目标案例的判刑类型，选择对应的神经网络模型
      // 并以目标案例刑期判决关键特征作为输入，推理具体刑期
      case '  $KFM_1$  ':  $TERM_{TC} = NNM_1(KKF_1)$ ;
      ....
      case '  $KFM_n$  ':  $TERM_{TC} = NNM_i(KKF_i)$ ;
    }

```

5.3.3 案例特征提取策略

在模型训练阶段，需要从案例库中的若干语义信息中提取案例关键特征，并在进行特征选择算法后训练决策树和神经网络模型；在模型的使用阶段，也需要在目标案例中提取相关特征进行判决结果的预测。关于案例特征的提取，从法官判案所考虑因素的角度考虑。以故意伤害罪为例，需提取案例中起因、经过、事后（结果及案情相关类内容）3类语义信息。

1) 起因

(1) 事出有因的程度（即被害人是否负有一定责任）；

(2) 激情犯罪的程度（若“伙同”、“纠集”、“被纠集”、“趁人不备”等则激情犯罪程度较低）。

2) 经过

- (1) 受害人数多少;
- (2) 作案工具攻击性强弱程度;
- (3) 攻击部位的要害程度;
- (4) 施暴方式的激烈程度;
- (5) 受案人伤情的严重程度。

3) 事后

- (1) 对受害人积极救治的程度;
- (2) 自首情节的程度;
- (3) 积极赔偿的程度 (是否达成民事赔偿调节协议等);
- (4) 得到被害人及其家属的谅解程度 (是否得到谅解、要求从轻等)。

基于刑事案件本体模型中的案情部分,在模型训练阶段需要从案例库中抽取相应的语义信息。需要对初步抽取得到的各案例特征的特征值进行分析,目的是发现案情差异性较大,各特征的取值差异很大,容易造成后期推理过程不收敛的情形;对于案例特征可以取多值的情况(如被告人可能对被害人身体的多个部位进行攻击),需对抽取的某些特征值进一步地整理、分类与规范化,使其满足使用决策树、神经网络等算法进行推理的要求。如表 5-1 所示,犯罪行为为事出有因的程度是通过案例特征“起因事件”表征的。根据特征值范围,将其进一步分为事因、酒后、其他原因 3 个子特征,其他原因又包含了矛盾后再次见面、被纠集、蓄意报复、无故等多个可能的取值。为了规范统计和便于推理,使用编号来表示特征值,案例特征详细处理情况见附录 E。

表 5-1 案例特征预处理

序号	案例号	事因	酒后	其他 原因	先施 暴方	其他 行为	打	...	伤情 等级	多人 作案
1	576_1	1	0	A8	B2	C4	1	...	D11	0
2	074_1	1	0	A4	B2	C4	0	...	D11	0
3	271_2	1	0	A8	B2	C2	1	...	D11	0
4	350_1	1	0	A8	B2	C4	0	...	D11	0

5.3.4 模型训练阶段的方案设计

在模型的训练阶段，通过历史案例，确定本体模型中案情类所描述的案例特征与判决结果之间的映射关系。按规则推理的粒度划分，将模型训练分为两个阶段：首先确定案例特征与所判刑罚类型之间的关系；再确定其与具体刑期之间的推理规则。

模型训练阶段的过程如图 5-8 所示。首先，基于特征选择算法，筛选与刑事判决类型紧密相关的案例特征；然后使用决策树算法，建立案件关键特征与刑事判决类型之间的推理规则；对每一类判决类型所对应的具体案件，训练神经网络模型，以案例各关键特征作为神经网络的输入层，以具体刑期作为神经网络输出，以便后期的使用过程中运用该神经网络模型推理出应判的刑期。

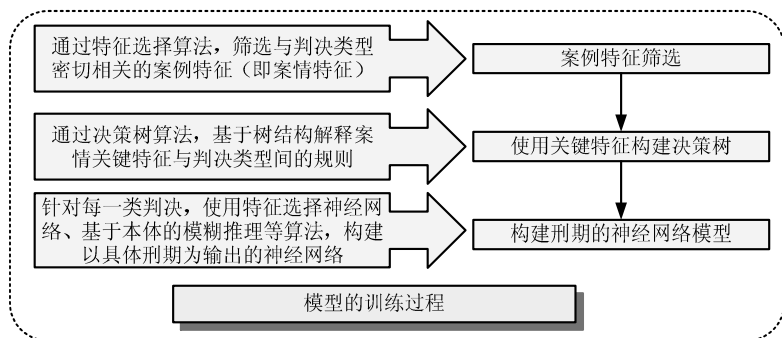


图 5-8 模型训练阶段的方案框架

5.3.5 模型使用阶段的方案设计

在推理模型的使用阶段，如图 5-9 所示：首先，抽取待检案例的案情特征，即基于刑事案件本体中案情类的相关属性抽取案情特征的特征值；然后，按照与模型训练阶段所提到的案例特征整理与分类方式相同的方法进行案例特征预处理；继而使用已完成训练的决策树和神经网络模型，先后预测待检案例应判决的刑罚类型和具体刑期；推理结束后，最终将根据法官对目标案例的量刑结果是否与推荐案例量刑结果相同，决定是否使新案例进入案例库。如果进入，则将案例信息使用本体描述格式表述，并放入案例库中。待新放入案例库中的案例到达一定的数量后，将重新进行特征选择、决策树和神经网络的训练，从而不断优化推理模型。这也是与我国实际审判工作与社会发展大背景，特别是与我国司法体系正趋于完善，量刑有动态变化相适应的。

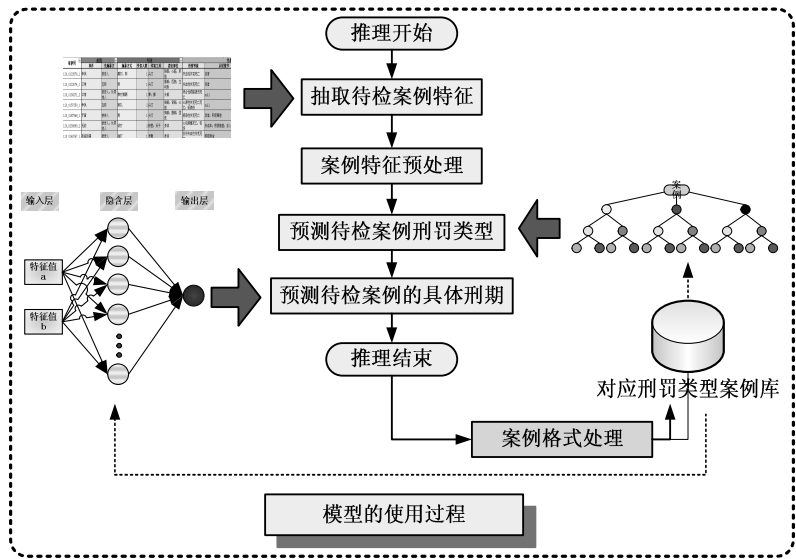


图 5-9 模型使用阶段的方案框架

5.4 刑事案件推理集成模型实现

在对法律推理规则挖掘的研究后，通过实验对上述过程设计和相关算法的可行性、有效性进行验证。

5.4.1 实验平台

本实验选用数据挖掘软件 Clementine 作为实验平台。

5.4.2 实验数据及预处理

1. 实验数据

实验数据采用××法院刑事案件中，一审判决后没有上诉或一审上诉二审维持的案件作为研究案例。在基于本体对刑事案件文档集进行标注抽取得到的案例库选取 150 个故意伤害罪案例，其中的 100 个为训练集，50 个为测试集，详细数据见附录 E。

2. 数据预处理

依本章前述的处理方法提取案例特征并进行以案例特征单值化为目的的预处理。共计 39 类案例特征，涵盖了犯罪行为的起因、经过、事后行为三方面语义。另外，将判决结果（即预测结果）分为刑、期、缓三部分，刑即为刑法类型，期即为具体刑期，缓即为缓刑期。其中，以故意伤害罪的法条为依据，将所判刑罚类型归纳为如下 7 类。

E1：三年以下有期徒刑。

E2：三年以下拘役。

E3：三年以下管制。

E4：三年以上十年以下有期徒刑。

E5：十年以上有期徒刑。

E6：无期徒刑。

E7：死刑。

将预处理后的案例语义特征数据集存储在 Excel 表格中，供模型建立使用，详见附录 E。针对每一个案例，使用 44 列语义特征值表示其特点，对于仅包含“有/无”两种可能取值的二值属性（如事因、酒后等），使用值 1 和 0 表示其有无；对于其他原因、先施暴方、其他行为、伤情等级、刑等多值属性，为其每一个可能的取值赋予一个字符值（即 B2 等标号），便于后续的数据挖掘操作。另外，在保留原始数据表的基础上，根据不同的判决类型，即特征属性“刑”的取值，将其整理至同一 Excel 工作簿的不同工作表中。

在原始判决书中抽取相应的语义信息，如表 5-2 所示。

表 5-2 原始判决书语义信息

序号	案例号	起 因		经 过				结 果	判决结果	
		事件	先施暴方	施暴方式	受害人数	作案工具	袭击部位	伤情等级	刑罚类型	具体刑期
1	576_1	争执	被告人	厮打；刺	1	尖刀	胸部；心脏；肝脏	失血性休克死亡	有期徒刑	15 年
2	074_1	口角	互殴	刺	1	尖刀	腹部；四肢；主动脉	失血性休克死亡	死刑	缓期 2 年

续表

序号	案例号	起 因		经 过				结 果	判决结果	
		事件	先施暴方	施暴方式	受害人数	作案工具	袭击部位	伤情等级	刑罚类型	具体刑期
3	271_2	口角	被告人；伙同他人	拳打脚踢	1	拳；脚	头部	闭合性颅脑损伤死亡	有期徒刑	15 年
4	350_1	争执	互殴	刺扎	2	尖刀	胸部；背部；心脏	心源性休克死亡；轻微伤	无期徒刑	
5	568_1	矛盾	被害人	刺	1	尖刀	胸部；腹部；四肢	感染性休克死亡	有期徒刑	12 年

通过对抽取得到的各案例特征的特征值分析，发现：由于各故意伤害案的犯罪过程（即案情）差异性较大，各特征的取值差异很大，容易造成后期推理过程不收敛的情形；对于案例特征可以取多值的情况（如被告人可能对被害人身体的多个部位进行攻击），也不利于使用决策树、神经网络等算法进行推理。因此，对抽取的某些特征值需要进行进一步的整理、分类与规范化。

1) 事件起因

将该部分可能的特征值总结如表 5-3 所示，表中第一行为一级特征；第二行为对一级特征的特征值的进一步分类，使在表述某一案例时，各二级特征有且只有一个特征值；第三行为各类中可能的特征值。规范特征值的目的是为了便于使用决策树、神经网络等数据挖掘算法，为各非二值特征属性的特征值赋予符号标识（如表中的 A1~A9）。二值特征属性的特征值规范为 0 或 1。

表 5-3 特征值总结-起因事件

一级特征	起因事件		
二级特征	事因	其他原因	酒后
特征值	是/否	矛盾后再次见面/被纠集/蓄意报复/无故/替他人出气/家庭矛盾	是/否
规范化后的特征值	1 / 0	A1（矛盾后再次见面）；A2（被纠集）；A3（蓄意报复）；A4（无故）；A5（替他人出气）；A6（家庭矛盾）；A7（其他）；A8（无）；A9（寻衅）	1 / 0

2) 先施暴方

一级特征中的先施暴方包括“单独”和“其他行为”两项，如表 5-4 所示。

表 5-4 特征值总结-先施暴方

一级特征	先施暴方	
二级特征	单独	其他行为
特征值	被害人/被告人/互殴	伙同他人/纠集他人/指使他人
规范化后的特征值	B1（被害人）；B2（被告人）； B3（互殴）	C1（伙同他人）；C2（纠集他人）； C3（指使他人）；C4（无）

3) 施暴方式

在总结该案情特征的特征值时，对多个可能的特征值进行归纳总结，总结为如表 5-5 所示的若干二级特征值，及其规范化后的特征值。特别注意，在实际案例中，有混合使用多种施暴方式的情况。

表 5-5 特征值总结-施暴方式

一级特征	施暴方式			
二级特征	打	刺	趁人不备	其他
特征值	是/否	是/否	是/否	是/否
规范化后的特征值	0 / 1	0 / 1	0 / 1	0 / 1

4) 作案工具

在总结该案情特征的特征值时，对多个可能的特征值进行归纳总结，总结为如表 5-6 所示的若干二级特征值，及其规范化后的特征值。特别注意，在实际案例中，有混合使用多种作案工具的情况。

表 5-6 特征值总结-作案工具

一级特征	作案工具			
二级特征	拳脚	钝器	锐器	其他

续表

特征值	是/否	是/否	是/否	是/否
规范化后的特征值	0 / 1	0 / 1	0 / 1	0 / 1

5) 袭击部位

在总结该案情特征的特征值时，对多个可能的特征值进行归纳总结，总结为如表 5-7 所示的若干二级特征值，及其规范化后的特征值。

表 5-7 特征值总结-袭击部位

一级特征	袭击部位							
二级特征	胸部	腹部	四肢	身体	头部	面部	眼部	颈部
特征值	是/否	是/否	是/否	是/否	是/否	是/否	是/否	是/否
规范化后的特征值	0 / 1	0 / 1	0 / 1	0 / 1	0 / 1	0 / 1	0 / 1	0 / 1

6) 伤情等级

该案情特征的特征值包括：轻微伤、轻伤（下限）、轻伤（偏轻）、轻伤、轻伤（偏重）、轻伤（上限）、重伤（下限）、重伤（偏轻）、重伤、重伤（偏重）、死亡，分别使用 D1~D11 标识各特征值。特别注意，若有多个受害人，则伤情等级按照伤情最重的。

7) 从轻情节

该案情特征的特征值包括：多人作案、被害人负有一定责任、未成年、限制刑事责任能力人、自首、积极救治、认罪、悔罪、积极赔偿、得到谅解、达成民事赔偿调解协议、被害人家属要求从轻，特征值取值均为：是/否。

8) 从重情节

该案情特征的特征值包括：无、累犯、主犯、劳教人员，特征值取值均为：是/否。

5.4.3 判决类型推理模型实验

如图 5-10 所示，为案例判决类型推断的决策树推理模型。图上半部分为模型的训练阶段：数据源为经过数据预处理的 100 组故意伤害罪案例，在“类型”节点中定义模型的输入、输出变量及其格式，后使用特征选择模块“案例特征选择”进行关键特征的筛选。上述部分运行后会产生模型“案例特征选择”（如图中的黄色菱形所示），然后通过该模型筛选特征，并训练后续的 CART 决策树模块，训练后即生成模型“刑类型_训练”。图下半部分为该模型的使用阶段：数据源为经过预处理的 50 组故意伤害罪案例训练集，同样在“类型”节点中定义模型的输入、输出变量及其格式，后使用“刑类型_训练”预测训练集的判决类型，将结果通过表格展示。

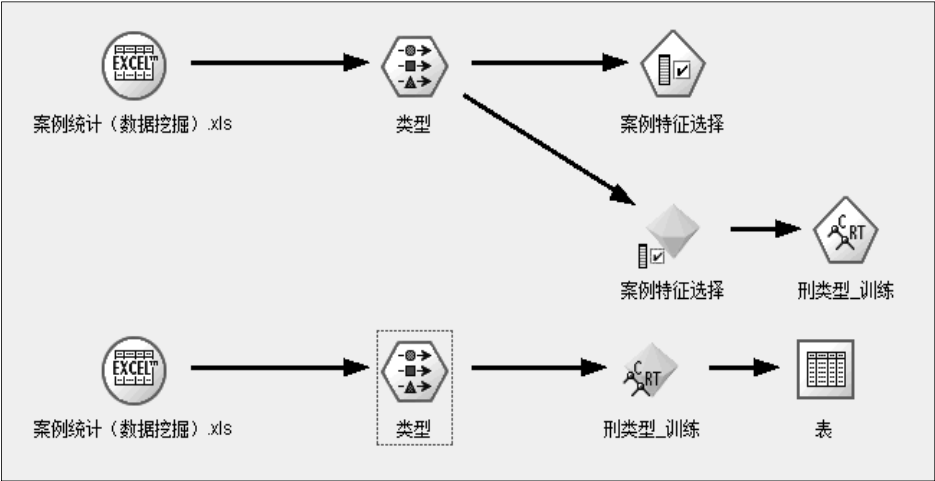


图 5-10 判决类型推理模型建立

1. 判决类型相关特征选择

通过类型工具定义模型的输入、输出变量及其数据格式。在该模型的建立中，序号、案例号、刑期、缓期为无关列，标示判决类型的列“刑”为输出，

其余列为模型的输入。对于二值属性，定义该字段的类型为“标识”；受害人数字段的类型定义为“范围”，其余多值属性的字段类型为“集”，即离散型。

特征选择模块的参数取值使用 Clementine 的默认取值，即将重要程度度量值（即前述的 importance 指数）的阈值设为 0.95，若取值超过阈值则认为该特征与预测过程相关性较大，若低于阈值则舍弃该特征。使用基于皮尔逊 χ^2 的方度量预测变量与目标变量之间的相关性。

筛选结果如图 5-11 所示，包括刺、打、腹部、面部、胸部、拳脚、锐器、伤情等级、受害人数、积极赔偿、认罪、自首共计 12 个特征，这些特征的重要程度度量值均在 0.95 以上。



图 5-11 特征选择结果

2. 判决类型决策树推理模型建立

基于筛选后的案例类型判决关键特征，建立决策树。定义 E1、E4、E5、E6 分别对应三年以下有期徒刑、三年以上十年以下有期徒刑、十年以上有期徒刑和无期徒刑。为了便于对决策树模型结果的理解，这里列出了实验中伤情等级的编号含义，如表 5-8 所示。

表 5-8 伤情等级在推理模型中的编号

伤情等级	编 号
轻微伤	D1
轻伤（下限）	D2
轻伤（偏轻）	D3
轻伤	D4
轻伤（偏重）	D5
轻伤（上限）	D6
重伤（下限）	D7
重伤（偏轻）	D8
重伤	D9
重伤（偏重）	D10
死亡	D11

决策树模型的训练结果如图 5-12 所示，树状图深度为 5。从本质上讲，决策树即为针对不同判决类型的规则集，决策树中自动生成的不同“模式”对应不同判决类型的规则。例如，在图 5-12 中，判决三年以下有期徒刑（E1）的特征规则集包括 3 种特征组合：

- （1）伤情等级为 D2、D4、D5 或 D6，该规则覆盖了 64 个案例。
- （2）伤情等级为 D7，且特征属性“胸部”的值为 0（即特征为假），且特征属性“打”的值为 1（即特征为真），且特征属性“面部”的值为 1，该规则属性覆盖了 1 个案例。

(3) 伤情等级为 D7, 且特征属性“胸部”的值为 1, 该规则覆盖了 1 个案例。

E1 规则集在测试集中的覆盖度为 100%。

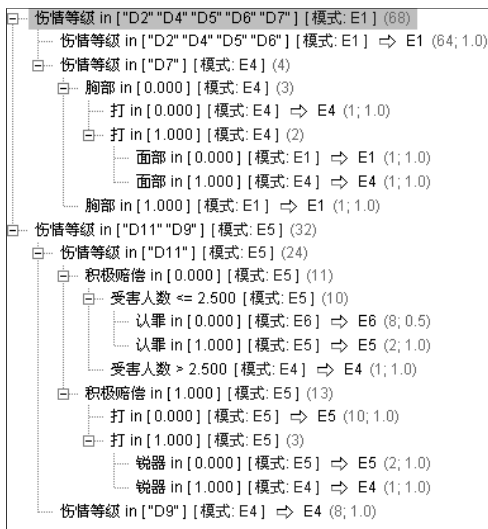


图 5-12 决策树模型训练结果

3. 判决类型规则推理模型验证

实验中选用的故意伤害罪案例是从事审判工作 10 年以上法官审理的一审判决后没有上诉或上诉后判决结果为维持原审判决的案件, 且这些案件没有在社会中产生舆情的案件, 这些案件具有可信性。实验中的正确标准是指模型推理结果与这些经验丰富的法官的判决结果相同。使用故意伤害罪的决策树模型对测试集中的案例判决刑罚类型进行推理预测, 正确率为 88%, 预测值的平均置信度为 0.82。

5.4.4 判决刑期推理模型实验

1. 判决刑期神经网络推理模型建立

对具体刑期推理模型的建立与训练过程与决策树模型训练阶段大致相似，不同之处在于：一是在数据源方面，由于每一刑罚类型都对应一个用于推理具体刑期的推理模型，因此模型训练阶段的数据源是按照不同的刑罚类型进行分类的；二是重新定义模型的输出项，将特征属性“期”、“缓”作为模型的输出；三是将数据推理模块由决策树模型换为神经网络模型，将神经网络方法选为“快速”。如图 5-13 所示，为在 Clementine 中建立的判决刑期推理模型，上半部分为模型的训练过程，下半部分为模型的测试过程，与前文的判决类型推理模型相似，这里不再详述。

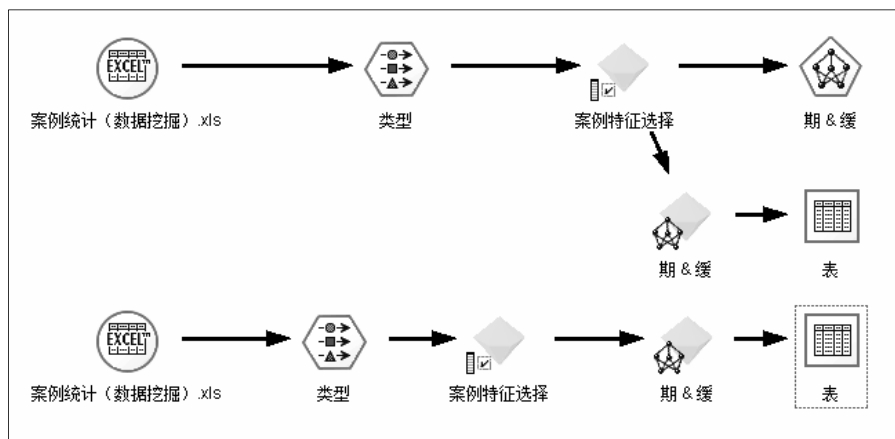


图 5-13 判决刑期推理模型建立

如图 5-14 所示，为判决刑期推理模型的训练结果，模型估计的准确性为 76.515%，包括一个含有 19 个神经元的输入层，一个含有 3 个神经元的隐藏层，一个含有 12 个神经元的输出层。图 5-14 的右图为 12 个输入变量的重要性值及排序。

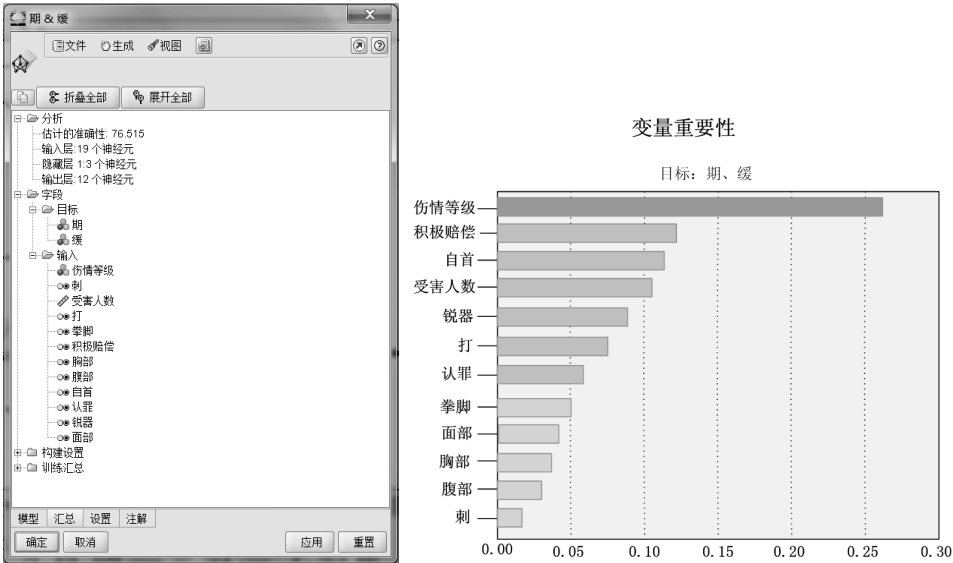


图 5-14 判决刑期推理模型训练结果

2. 判决刑期神经网络推理模型验证

实验中，仍然选择了具有丰富经验的法官审理的没有争议的近年的案件作为实验对象，以确保实验的可信性。在实验中，以“三年以下有期徒刑（E1）”为例，训练集共计 66 个案例，测试集 37 个案例。刑期和缓刑期两个预测结果完全匹配的准确率为 29.73%。考虑到法官自由裁量权及预测具体刑期的难度和复杂性，通过咨询法律专家，可将正负 0.5 年作为可以接受的误差范围，则推理结果的准确率为 72.97%。

5.5 刑事案件推理集成模型评价

本书分别通过测试集对规则与案例推理集成决策模型进行了实验评价，结果如表 5-9 所示，为实验中对判决类型、判决刑期两个实验的结果整理。总

体而言，经过实验验证，基于本体的刑事案件案例推理集成模型可以较好地预测目标，对刑罚类型推理结果的准确度较高；对具体刑期和缓刑期的推理结果的准确度虽有待进一步提高，但预测值与真实值的偏离量较小，考虑多案例推理逻辑的复杂性，并征求审判实务专家意见，认为这一推理结果对审判人员的科学量刑具有很大的支持作用，对司法标准的统一有积极作用，具有创新应用价值。

表 5-9 实验结果

推理过程	使用算法	训练集案例数量	测试集案例数量	可允许的误差范围（年）	准确率%
推理判决类型	特征选择+决策树	100	50		88
推理具体刑期	特征选择+神经网络	66	37	0 年	29.73
推理具体刑期	特征选择+神经网络	66	37	0.5 年	72.97

5.6 本章小结

本章以刑事案件量刑决策需求为出发点，针对司法审判领域多变量和动态变化决策的特点，提出了案例库支持下的规则推理和案例推理集成决策模型。首先，本章研究了基于刑事案件本体的决策推理方案，从基于本体的案例特征提取、推理模型的训练、使用三个阶段构建了方案框架。将决策目标分成决策结果类型和决策量化指标，即刑罚类型和刑期，使决策过程可以采用更具有针对性的推理方法，解决决策中的问题。其次，基于特征选择、决策树模型、神经网络模型等数据挖掘算法实现了规则推理和案例推理的集成，二者以规则推理出的刑罚类型与对应刑罚类型的神经网络模型进行“级联”，按刑罚类型对应的神经网络模型对刑期进行推理，这种结合拓宽了决策管理理论与方法的应用，

具有一定的应用和学术价值。第三，经过实验验证，本方案有效地解决了判决类型、判决刑期的辅助决策，准确率达到 72.97%，进一步证明了刑事案件推理集成决策模型的合理性。研究发现，这种推理决策使选择的训练案例集针对性更强，训练得到的神经网络模型针对刑罚类型颗粒度更小，较之以前仅通过案件特征相似度计算选择相似案例的做法有了很大进步。

刑事案件量刑决策辅助系统原型设计

基于信息化支持法官量刑决策必须是可视化、自动化的决策推理过程，需要将基于本体的半结构决策数据结构化、基于 GA-KNN 的案例属性优化与检索和基于规则和案例推理集成的决策机制的研究成果实例化。本章研究问题是构建刑事案件量刑决策辅助系统原型，解决支持量刑决策机制与信息技术结合问题。研究思路是以系统构建基本模型为基础，从系统需求、系统设计、关键算法实现和展示等几个层面进行前瞻性总体设计，通过实验验证原型的可行性。最终实现为决策辅助系统开发提供原型推荐的目的，为基于规则和案例推理集成的刑事案件量刑决策机制研究成果的应用提供可行性建设方案。

6.1 需求分析

首先对刑事案件量刑决策系统原型进行需求分析，从系统需求概述和功能设计的角度进行分析。

6.1.1 系统需求概述

设计基于本体案例库的刑事案例判决系统原型的目的是，应用第5章中的基于本体的刑事案件推理方案，设计系统原型，以实现面向海量数据、自动化的决策推理系统，并可实际应用于法院工作人员在刑事案件过程中的判决结果推理、相似案例推荐等过程，为其工作的进行提供良好的数据支持、决策支撑。将系统建立的目标分解为如下三点需求：

- (1) 刑事案件本体案例库的自动化建立。
- (2) 基于本体案例库中的案例，建立训练集，根据案例的类别，自动化地完成推理模型的训练过程。
- (3) 对于已经训练好的推理模型，可以对目标案例的判决类型、刑期、相似案例等进行推理，并向法官推送推理结果。

6.1.2 系统功能需求

结合系统需求并对其进行进一步细化，设计了本系统的功能框架，如图6-1所示。将系统分为两个子系统：案例库构建子系统与案例推理与推荐子系统。案例库构建子系统包含本体案例库构建模块，主要有三个功能：一是为系统导入

OWL 格式的本体框架文件，即本体的结构模型；二是将批量的历史案例导入本体案例库的功能；三是提供案例本体知识库的检索功能。案例推理与推荐子系统包含案例推理模块和案例推荐模块。其中，案例推理模块主要有三个功能：一是经案例库中的案例数据预处理，使其便于后期的推理过程；二是使用案例库中的历史案例构成训练集，训练推理模型；三是使用训练好的模型，针对目标案例，进行基于本体案例库的案例推理。案例推荐模块与案例推理模块有一定的耦合性，主要有两个功能：一是基于推理的结果进行相似案例的推荐；二是针对已判决的案件，提供相似案例的推荐，该功能可以独立于案例推理功能存在。

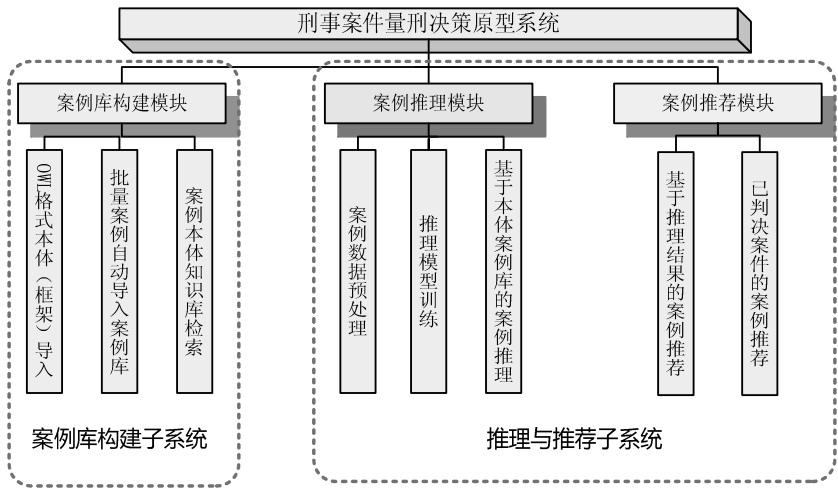


图 6-1 系统功能框架

6.2 系统总体设计

本节中，对刑事案件量刑决策辅助系统原型的工作流程、系统角色模型、系统整体架构等进行梳理和介绍。

6.2.1 系统工作流程

从任务完成的角度看,系统工作流程如图 6-2 所示。首先需要完成刑事案件原始本体案例库的建立:基于设计好的刑事案件本体框架,对批量的刑事案件案例文本进行案例语义标注与特征抽取,将非结构化的海量文本信息转化为富含语义信息且结构化程度较好的本体案例库,并提供对案例本体案例库的检索。在案例推理与推荐子系统中,先将刑事案件原始本体案例库中的信息进行预处理,面向案例推理过程,建立刑事案件推理本体案例库,该本体案例库着重关注与案件审判相关的语义信息的表述方式。利用刑事案件推理本体案例库,实现对推理模型的建立与训练,继而对法官输入的目标案例进行刑期的推理和相似案例的推荐。图中黄色矩形表示法官直接驱动的系统流程,包括案例本体案例库的检索、目标案例信息输入、目标案例推理、基于推理结果的案例推荐和已判决案件的案例推荐。其他环节相当于数据准备阶段的工作,均可在系统建立初期由后台人员完成。系统运行期间,也需要基于法官使用过程中输入的目标案例和已判决案例,不断丰富刑事案件原始本体案例库和刑事案件推理本体案例库,并周期地优化推理模型,使模型更加符合案例的实际特点。下面从系统所涉及角色及其操作的角度,对工作流程进行较为细致的划分。

从角色模型的角度看,分为两部分:一是在系统搭建初期,主要针对系统管理员的推理模型训练阶段,如图 6-3 所示。二是在系统的使用阶段,主要针对法官(即审判人员)的推理模型使用阶段,如图 6-4 所示。图中所涉及的角色除系统管理员、法官(审判人员)之外,针对本系统内部设计的一些不需法官操作直接驱动的关键步骤与算法,还增加了对系统自动化流程的展示。对图中的具体流程解释与对图 6-2 的解析相同,只是从行为对象的层面加以强调,这里不再赘述。特别地,图 6-4 中,在法官使用的过程中,系统管理员提供技

术咨询与支持服务，并周期进行推理模型的参数调整、重新训练等工作，以优化推理模型，使系统性能达到令法官满意的水平。

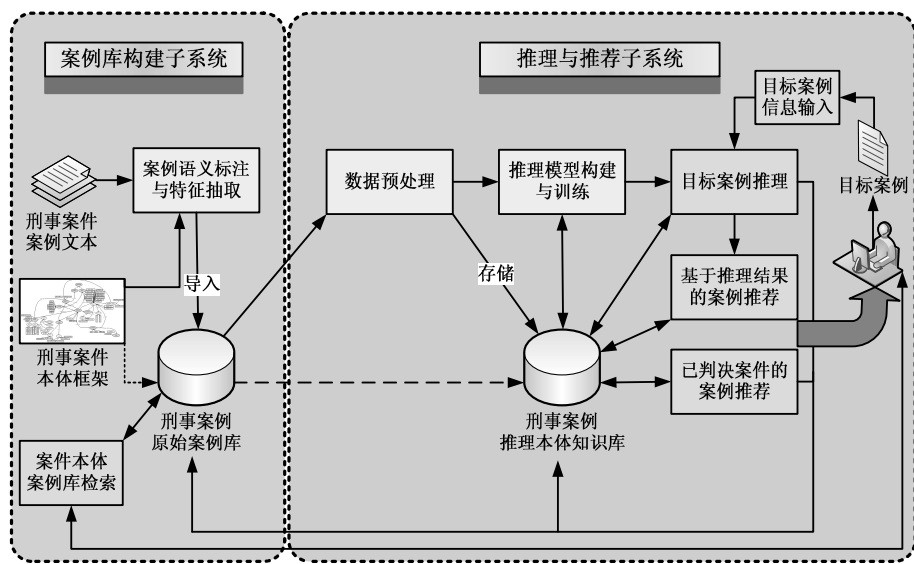


图 6-2 系统工作流程

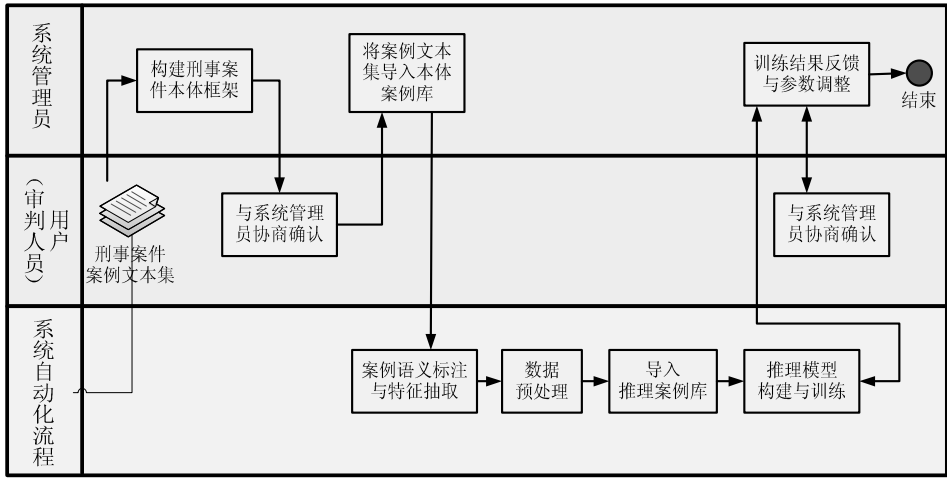


图 6-3 模型训练阶段角色模型

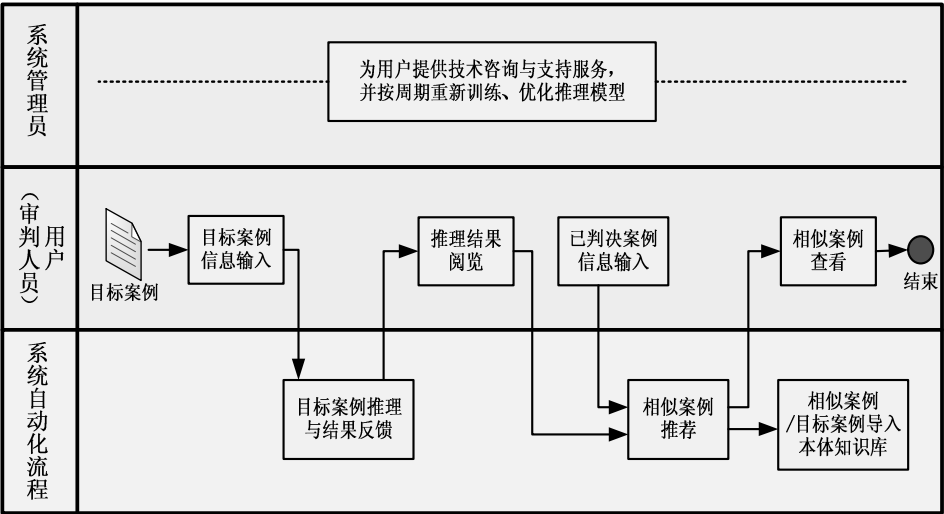


图 6-4 模型使用阶段角色模型

6.2.2 系统架构设计

如图 6-5 所示，系统架构由底向上分为硬件资源层、软件平台层、核心组件层、系统功能层和用户交互层。其中，硬件资源层是系统搭建的硬件基础，包括主机、网络布线、存储服务器搭建等。在硬件资源的基础上，搭建软件平台层，包括各类开发平台及关键技术。核心组件层是对于系统关键功能的抽象，软件的功能由组件构成，在升级时，只需要对组件进行调整，不需重构功能。系统功能层为各个可以交付给法官的工具包，法官可以用其实现相应功能。最终，系统可以通过 Web 客户端、PC 客户端、移动设备客户端等设计界面方式实现与法官的交互。

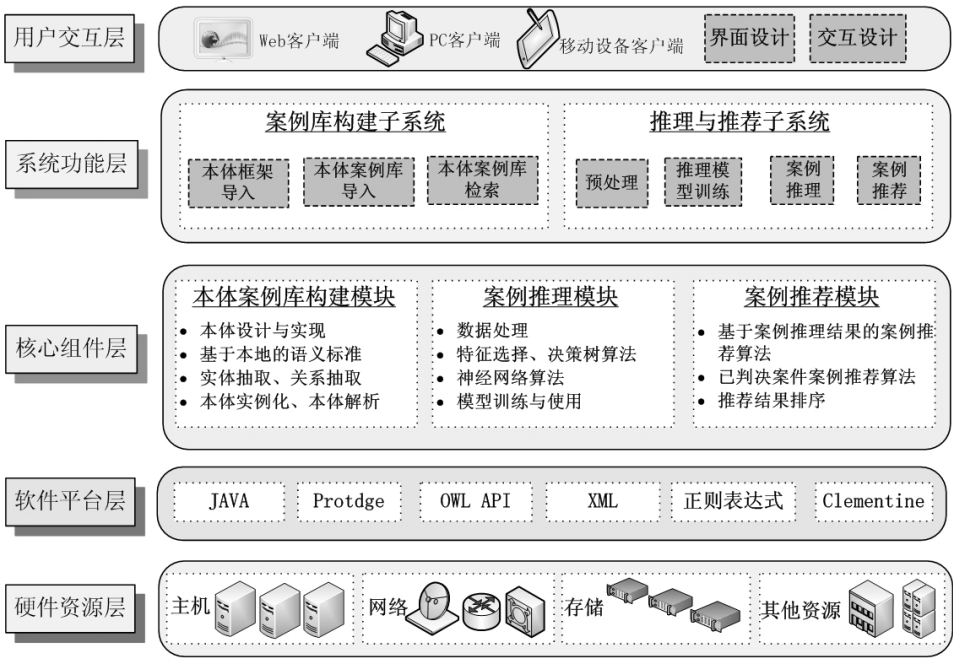


图 6-5 系统架构设计

6.3 功能模块实现

分别介绍本体案例库构建模块、案例推理模块及案例推荐模块的设计细节和相关算法。

6.3.1 案例库模块实现

按照第 3 章中的刑事案件本体设计与建立方法，在 Protégé 中进行刑事案件本体框架的设计与实现，如图 6-6 所示，在 Protégé 中对本体类、对象数据与数据属性进行编辑，最后使用后缀为 OWL 的本体文件格式保存刑事案件本体框架。

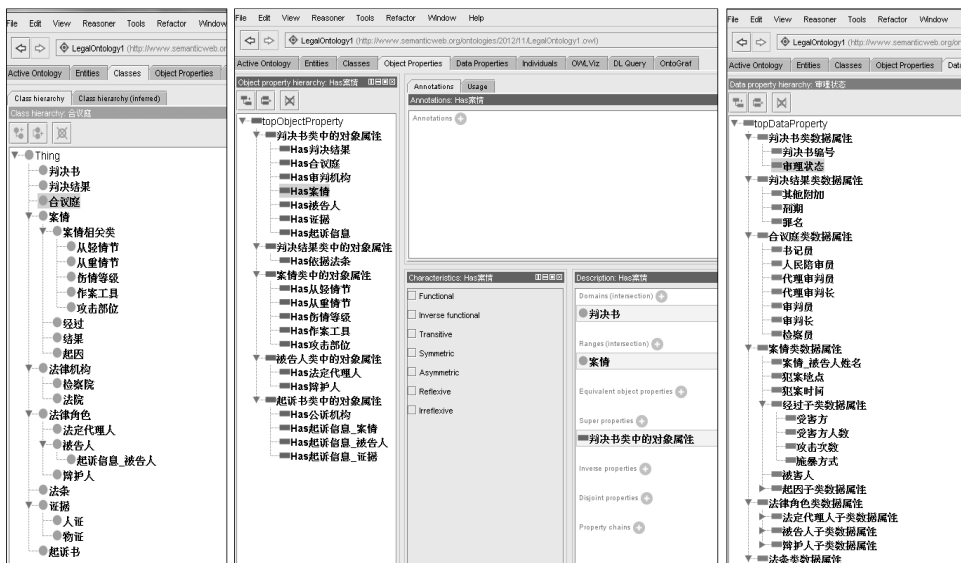


图 6-6 Protégé 中设计与实现本体框架

6.3.2 决策推理模块实现

在案例推理模块中，首先考虑的是对存储于本体案例库中的案例原始信息进行预处理，其目的是便于在推理模型的建立与训练、目标案例推理等步骤中调用数据挖掘算法进行推理。

1. 推理模型建立与训练算法

使用刑事案例推理本体案例库中的案例作为推理模型的训练集，使用本书第5章所述的算法进行基于特征选择、决策树、神经网络等数据挖掘算法的推理模型的建立与训练，这里不再赘述。算法流程图如6-7所示。

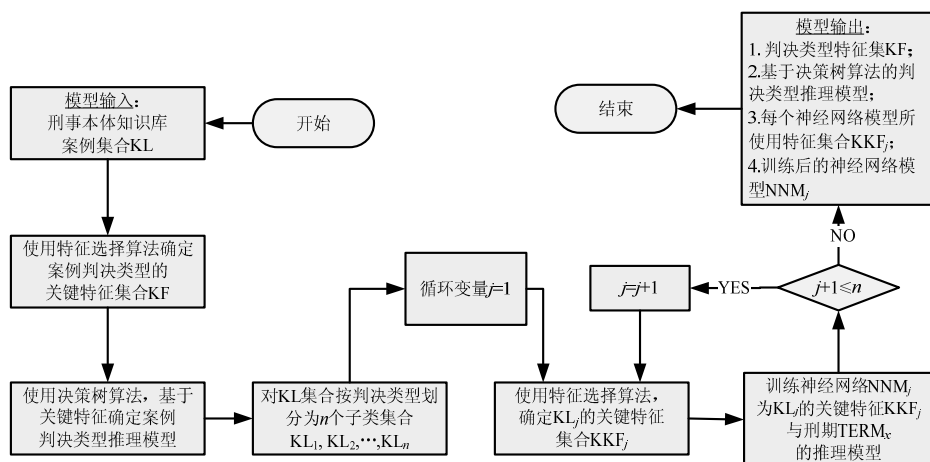


图 6-7 推理模型训练算法流程图

算法伪代码如下：

算法输入：刑事案例推理本体案例库 KL。

算法输出：决策树模型使用的特征集合 KF；

训练后的决策树模型 DT（即判决类型 KL_i 的关键特征模式 KFM_i ）；

每个神经网络模型所使用特征集合 KKF_j ；

训练后的神经网络模型 NNM_j

（即 KL_j 的关键特征集合 KKF_j 与刑期 $TERM_x$ 的推理模型）。

KL=刑事案例推理本体案例库；

$KF = FD(KL)$ ；//使用特征选择算法 FD 确定 KF 为案例库 KL 中案例的关键特征集合

KL_1, KL_2, \dots, KL_n 为 n 个判决类型；//即为对 KL 的 n 个划分

for(i = 1; i <= n; i++)

{

$KFM_i = DT(KF, KL_i)$ //使用决策树确定判决类型 KL_i 的关键特征模式为 KFM_i

}

for(j = 1; j <= n; j++)

{

```
KKFj = FD(KLj) ; // 再次使用特征选择算法 FD 确定 KLj 的关键特征集合
为 KKFj

// 训练神经网络模型 NNMj 为 KLj 的关键特征 KKFj 与刑期 TERMx 的推理
模型

NNMj = NN(KKFj, TERMx) ;

}
```

2. 目标案例推理算法

使用本书第 5 章所述的算法进行目标案例推理, 算法流程图如图 6-8 所示。

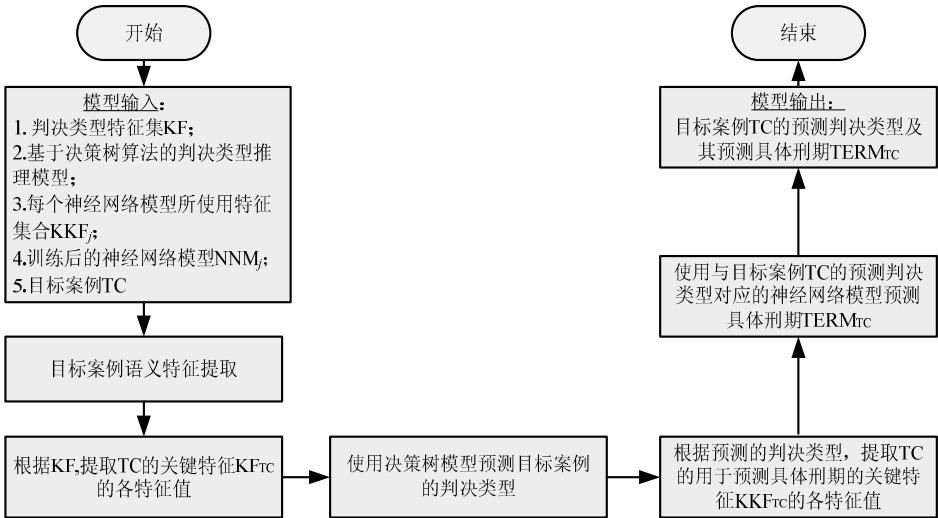


图 6-8 集成推理模型算法流程图

算法伪代码如下：

```
算法输入：决策树模型使用的特征集合 KF；
训练后的决策树模型 DT（即判决类型 KLi 的关键特征模式 KFMi）；
        每个神经网络模型所使用特征集合 KKFj；
        训练后的神经网络模型 NNMj
        （即 KLj 的关键特征集合 KKFj 与刑期 TERMx 的推理模型）。
算法输出：目标案例的参考刑期 TERMTC
```

```

目标案例语义特征提取；
根据模型训练阶段的结果选取关键特征  $KF_{TC}$ ；
//使用决策树确定的关键特征集合  $KFM_i$  与判决类型  $KL_i$  的映射关系，
//确定目标案例的判决类型。

switch( $KFM_{TC}$ )
{ //针对目标案例的判决类型  $KL_{TC}$ ，选择对应的神经网络模型  $NNM_i$ 
  //并以目标案例关键特征集合  $KKF_i$  作为输入，推理具体刑期  $TERM_{TC}$ 
  case '  $KFM_1$  ': {
       $KL_{TC} = 1$ ；
       $TERM_{TC} = NNM_1(KKF_1)$ ；
  }
  .....
  case '  $KFM_i$  ': {
       $KL_{TC} = i$ ；
       $TERM_{TC} = NNM_i(KKF_i)$ ；
  }
}
    
```

6.3.3 案例推荐模块实现

案例推荐算法基本思路是，在与目标案例推断的判决类型相对应的案例集合中，在一定范围内，把近似符合刑期和刑期关键特征的案例推荐给法官。伪代码如下所示。

```

算法输入：目标案例 TC；
           TC 判决类型  $KL_{TC}$ ；
           TC 推理刑期  $TERM_{TC}$ ；
算法输出：目标案例 TC 的相似案例
            $KL_{TC} = KL_i$ ；//其推理得到的判决类型为 i
    
```

```

TERMTC = x; // 其推理得到的刑期为 x
寻找相似案例的数据集合为 KLTC 判决类型的所有案例, SimL[ ];
for(p = 0; p < SimL.size(); p++){
if (SimL[ ] 中第 p 个案例的刑期与 TC 的刑期相差正负 0.5 年以内)
    将 p 加入刑期匹配集 SetTM[ ];
if (SimL[ ] 中第 p 个案例的特征集 KKFi 的取值与 TC 的取值 80 以上相同)
    将 p 加入特征值匹配集 SetFM[ ];
}
//输出相似案例
if(SetTM&SetFM !=null) print(SetTM&SetFM);
else if(SetFM!=null) print(SetFM);
else if(SetTM!=null) print(SetTM);
else 减少刑期或特征值的匹配度, 再次进行相似案例的计算;

```

6.4 系统原型界面设计

案例库建立模块的界面设计如图 6-9 所示, 该模块界面在系统主界面的“知识库构建”选项卡中呈现。界面的上半部分为刑事审判本体框架 (即 OWL 文件)、审判文档文件夹、知识库文件的浏览和选择, 单击“确定”按钮后验证上述文件路径的可用性, 三者均可用后, 即可单击“导入”按钮完成审判案例文档的批量语义标注、实体抽取并导入本体框架中, 形成本体案例库。其中, 案例库文件可以是已经存在的本体案例库文件, 也可以采用新建文件的形式。界面的下半部分为刑事审判案例本体知识库的检索界面, 在左侧的文本框中输入基于 OWL 语言的本体查询检索表达式, 出于易用性的考虑, 法官可以单击“常用检索式”按钮获取一些较为常用的检索式, 并通过“帮助”按钮提供详细表达式书写帮助文档。单击“查询”按钮后, 检索结果将在右侧的文本框中显示。



图 6-9 案例库建立模块界面设计

在本体案例库建立模块完成原始本体案例库建立之后，即在程序后台进行数据预处理、推理模型建立与训练过程。上述过程不需法官操作，因此未在界面上提供相关操作接口。案例推理部分界面设计如图 6-10 所示，法官只需完成案例库文件的选择和目标案例的输入，即可得到推理结果，即推理得到的案件判决结果。



图 6-10 案例推理界面设计

其中，目标案例的语义特征提取使用法官输入的方式实现，单击“目标案例输入”按钮即为目标案例信息输入界面，如图 6-11 至图 6-13 所示。

目标案例信息输入

案例编号：

案例起因

案例经过

案例结果

1. 起因事件

是否具有事由：

☐ 是

☐ 否

是否是酒后行为：

☐ 是

☐ 否

其他原因：

A1 矛盾后再次见面

2. 先施暴方

先施暴方：

B1 被害人

其他行为：

C1 伙同他人

保存

图 6-11 案例起因信息输入界面

目标案例信息输入

案例编号：

案例起因

案例经过

案例结果

3. 施暴方式

☐ 打

☐ 刺

☐ 趁人不备

☐ 其他

4. 受害人数：

0

5. 作案工具

☐ 拳脚

☐ 钝器

☐ 锐器

☐ 其他

6. 袭击部位

☐ 胸部

☐ 腹部

☐ 四肢

☐ 身体

☐ 头部

☐ 面部

☐ 眼部

☐ 颈部

保存

图 6-12 案例经过信息输入界面

目标案例信息输入

案例编号：

案例起因

案例经过

案例结果

7. 伤情等级：

D1 轻微伤

8. 从轻条件

☐ 多人作案

☐ 被害人负有一定责任

☐ 未成年

☐ 限制刑事责任能力人

☐ 自首

☐ 积极救治

☐ 认罪

☐ 悔罪

☐ 积极赔偿

☐ 得到谅解

☐ 达成民事赔偿协议

☐ 被害人方面要求从轻

☐ 其他

9. 从重条件

☐ 累犯

☐ 主犯

☐ 劳教人员

保存

图 6-13 案例结果信息输入界面

• 167 •

如图 6-14 所示,将案例推理与案例推荐的法官交互部分在同一界面上布局,需要先完成目标案例推理,才能单击“目标案例推荐”按钮,即在右边的界面中列出相似案例。选中某一案例后,单击“查看详细信息”按钮即可查看该案例的详细文档。单击“已判决案例推荐”按钮,输入已判决案例的相关信息(输入界面与目标案例信息输入界面相似,只增加判决类型、具体刑期信息的输入),即可对历史案例进行相似案例的推荐。

系统会根据法官的量刑结果,自动判断是否加入新案例,更新案例库,对推理模型进行重新训练。模型的训练以年为周期为宜。



图 6-14 案例推理与推荐界面

6.5 本章小结

本章从系统需求和系统设计出发,构建了刑事案件量刑决策辅助系统原型,实现了刑事案例库构建、审判案例推理和审判案例推荐三大系统的建立。通过

基于本体的刑事案件信息抽取模板,实现了判决书文档集的自动规则化;通过目标案例特征的人机互动输入,实现了决策推理系统输入,并在最优案例库和模型库支持下,实现推理集成决策机制和推理“刑”、“期”和推荐案例的可视化。系统原型设计和11 000个案例支持推理,在我国司法审判领域首次实现了案件知识获取、决策推理和系统应用三大功能的集成,满足了刑事案件量刑决策中对决策信息规则化、推理过程自动化和推荐结果可视化的要求。

结论与展望

本书针对司法领域量刑判决中的量刑尺度的统一性问题，基于司法领域的海量司法案例，重点研究法律文本的结构化、案例的科学检索、规则与案例集成的量刑决策。本书运用本体方法、GA-KNN 结合检索、规则与案例集成决策等方法，构建了基于本体的刑事案例表示方法，提出和论证了对于多属性、多案例、多要素背景下，基于 GA-KNN 的案例属性优化与检索针对性方案，揭示了基于规则与案例推理集成的量刑决策机制。以 11 000 个故意伤害罪的半结构化案例文档集为初始样本，通过本体技术进行决策信息规则化、结构化，采用 GA-KNN 算法进行属性优化与案例检索，采用决策树和神经网络模型构建推理集成决策模型，通过与实际量刑决策结果比较论证了该方法对支持量刑决策的意义。

本书的主要研究结论及研究理论与应用意义体现在以下方面：

第一，运用本体理论规范和领域知识表达，通过建立刑事案件本体，可以实现以罪名为单位，对刑事案件相关的概念、类和属性进行规范化表达；同时可以实现半结构化文本决策数据的自动抽取，实现刑事案件实体信息的自动化描述性抽取；也可以通过数据仓库技术，对刑事案件决策信息进行逻辑设计，将描述性实体信息规则化，并形成规则化案例库。本书研究结论以本体为核心，

从模型、实验和评价三个方面实现了司法案例从半结构化数据向结构化案例库的转化。提供了更多决策信息可利用的重要途径，突破了从半结构化数据中获取决策信息的瓶颈，首次实现了本体理论在司法案例领域量刑决策中的扩展性应用。

第二，司法领域的多案例、多属性、多要素特征为推理机制依赖的案例库满足快速检索及适时性提出了新的要求。刑事案例需要确立案例推理精确的目标与案例检索环节的合理路径；将 GA 和 KNN 方法进行组合，用于刑事案例的检索，提高了多案例、多属性、多要素司法审判案例应用的检索精度及检索速度。研究结论不仅满足了刑事案例检索精确度及检索效率要求，而且拓宽了管理科学与信息技术交叉研究的范围，实现了决策过程依赖特定决策基础信息的最优集合。

第三，决策树模型和神经网络模型的组合模式，可以实现法规与案例在决策推理中有机结合，形成刑事案件量刑决策中规则推理和案例推理集成决策机制，实现刑事案例量刑决策的科学化和规范化。这一模式是未来司法量刑决策的可靠依据。这一研究利用数据挖掘方法对领域知识进行深度挖掘，可以找到决策中潜在的关键信息，如量刑决策特征中的暴力方式细节与决策结果的关系、量刑结果类型与各犯罪特征集的对应等。研究结论对于两种推理的集成充实了管理决策方法，同时，拓宽了领域知识的挖掘和利用机制。

本书是在我国遵循成文法量刑的基础上，对充分利用案例支持法官有效约束自由裁量权，使量刑决策结果趋于统一化问题进行的前瞻性研究，其研究成果除了在司法审判领域使用外，也适用于具有半结构化大数据，需要规则与案例并重支持的领域，如医疗领域等。

尽管本研究取得了一些阶段性成果，但仍存在一些局限性有待在以后的研究中解决和深入。主要局限性体现在：

第一，从法律关系角度来说，法律领域的法律关系复杂性比研究中遇到的情况更多，本书针对刑事案件量刑决策进行研究，是探索性、前瞻性研究，是对支持法官决策方法的突破，但要真正实现全面支持司法审判工作，仍然还有许多法律关系需要厘清，如刑事案件中的缓刑、刑事附带民事、社会舆论对决策的影响等。

第二，从数据挖掘技术角度来说，需要继续对推理模型的算法进行优化，结合本体模型的特征，进一步加强基于“证据”的推理研究，使其结果更加准确。同时，进一步对案例推理原型系统进行改进，从面向法官用户的角度分析，将本体推理问题转化为随时供法官应用的信息检索问题，提高系统的可用性和易用性。

第三，从研究基础数据角度来说，选用的是特定区域法院的故意伤害罪案件，其虽然具有一定的代表性，但对于支持全国范围的量刑决策系统前瞻性研究，还需要进一步选择更多区域、更多刑事案件类型的案例进行验证，使研究结论更具普适性。

此外，案例库仅在本书研究中得到部分应用，它还可以用在针对法律关键事件的关联关系、审判规律、审判态势等问题的研究中，使其得到充分利用。

刑事案件本体类与属性的详细设计

类：法律机构

子类：法院

子类：检察院

类：起诉书

数据属性：起诉书编号

数据属性：提起公诉日期

对象属性：Has 起诉信息_被告人（类：起诉书 $\mathbf{\hat{a}}$ 类：起诉信息_被告人）

对象属性：Has 公诉机关（类：起诉书 $\mathbf{\hat{a}}$ 类：检察院）

对象属性：Has 起诉信息_案情（类：起诉书 $\mathbf{\hat{a}}$ 类：案情）

对象属性: Has 起诉信息_证据 (类: 起诉书 \rightarrow 类: 证据)

类: 合议庭

数据属性: 审判长

数据属性: 审判员

数据属性: 代理审判员

数据属性: 代理审判员

数据属性: 人民陪审员

数据属性: 书记员

数据属性: 检察员

对象属性: Has 组织法院 (类: 合议庭 \rightarrow 类: 法院)

类: 法律角色

子类: 被告人

数据属性: 被告人姓名

数据属性: 被告人性别

数据属性: 被告人年龄

数据属性: 被告人出生年月

数据属性: 被告人民族

数据属性：被告人住址

数据属性：被告人户籍地

数据属性：被告人文化水平

数据属性：被告人职业

数据属性：涉嫌罪名

数据属性：被羁押日期

数据属性：被逮捕日期

数据属性：现羁押地

数据属性：是否累犯

对象属性：Has 法定代理人（类：被告人 **a** 类： 法定代理人）

对象属性：Has 辩护人（类：被告人 **a** 类： 辩护人 ）

子类：起诉信息_被告人

数据属性：起诉信息_被告人姓名

数据属性：起诉信息_指控罪名

子类：辩护人

数据属性：辩护人姓名

数据属性：辩护人单位

子类：法定代理人

数据属性：法定代理人姓名

数据属性：法定代理人年龄

数据属性：法定代理人出生日期

数据属性：法定代理人民族

数据属性：法定代理人职业

数据属性：法定代理人住址

数据属性：与被告人关系

类：案情

数据属性：案情_被告人姓名

数据属性：犯案时间

数据属性：犯案地点

数据属性：被害人

子类：起因

数据属性：事件

数据属性：先施暴方

子类：经过

数据属性：施暴方式

数据属性：受害方

数据属性：受害人个数

数据属性：攻击次数

对象属性：Has 攻击部位（类：经过 **à** 类：攻击部位）

对象属性：Has 作案工具（类：经过 **à** 类：作案工具）

子类：结果

对象属性：Has 伤情等级（类：结果 **à** 类：伤情等级）

对象属性：Has 从轻情节（类：结果 **à** 类：从轻情节）

对象属性：Has 从重情节（类：结果 **à** 类：从重情节）

子类：案情相关类

子类：攻击部位

子类：作案工具

子类：伤情等级

子类：从轻情节

子类：从重情节

类：证据

子类：物证

数据属性：获取物证时间

数据属性：归属

数据属性：物品

子类：人证

数据属性：证人证明时间

数据属性：证人涉案关系

数据属性：证人证明地点

数据属性：证人证明内容

类：判决结果

数据属性：罪名

数据属性：刑期

数据属性：其他附加

对象属性：Has 依据法条（类：判决结果 **a** 类：法条）

类：法条

数据属性：法条编号

数据属性：法条内容

类：判决书

数据属性：审理状态

数据属性：判决书编号

对象属性：Has 起诉信息（类：判决书 à 类：起诉书）

对象属性：Has 审判机构（类：判决书 à 类：法院）

对象属性：Has 合议庭（类：判决书 à 类：合议庭）

对象属性：Has 被告人（类：判决书 à 类：被告人）

对象属性：Has 案情（类：判决书 à 类：案情）

对象属性：Has 证据（类：判决书 à 类：证据）

对象属性：Has 判决结果（类：判决书 à 类：判决结果）

刑事案件本体 OWL 文件

```
<?xml version="1.0"?>
<!DOCTYPE rdf:RDF [
    <!ENTITY owl "http://www.w3.org/2002/07/owl#" >
    <!ENTITY xsd "http://www.w3.org/2001/XMLSchema#" >
    <!ENTITY rdfs "http://www.w3.org/2000/01/rdf-schema#" >
    <!ENTITY rdf "http://www.w3.org/1999/02/22-rdf-syntax-
ns#" >
    <!ENTITY LegalOntology1 "http://www.semanticweb.org/
ontologies/2012/11/LegalOntology1.owl#" >
]>

<rdf:RDF xmlns="http://www.w3.org/2002/07/owl#"
    xml:base="http://www.w3.org/2002/07/owl"
    xmlns:LegalOntology1="http://www.semanticweb.org/ontologies/
2012/11/LegalOntology1.owl#"
    xmlns:xsd="http://www.w3.org/2001/XMLSchema#"
    xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#"
    xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
    xmlns:owl="http://www.w3.org/2002/07/owl#">
```

```
<Ontology rdf:about="http://www.semanticweb.org/ontologies/
2012/11/LegalOntology1.owl"/>
```

```
<!--
```

```
////////////////////////////////////
```

```
//
```

```
// Object Properties
```

```
//
```

```
////////////////////////////////////
```

```
-->
```

```
<!-- http://www.semanticweb.org/ontologies/2012/11/Legal
Ontology1.owl#Has 从轻情节 -->
```

```
<ObjectProperty rdf:about="&LegalOntology1;Has 从轻情节">
```

```
<rdfs:range rdf:resource="&LegalOntology1;从轻情节"/>
```

```
<rdfs:subPropertyOf rdf:resource="&LegalOntology1;案
情类中的对象属性"/>
```

```
<rdfs:domain rdf:resource="&LegalOntology1;结果"/>
```

```
</ObjectProperty>
```

```
<!-- http://www.semanticweb.org/ontologies/2012/11/Legal
Ontology1.owl#Has 从重情节 -->
```

```
<ObjectProperty rdf:about="&LegalOntology1;Has 从重情节">
```

```
<rdfs:range rdf:resource="&LegalOntology1;从重情节"/>
```

```
<rdfs:subPropertyOf rdf:resource="&LegalOntology1;案
情类中的对象属性"/>
```

```
<rdfs:domain rdf:resource="&LegalOntology1;结果"/>
```

```
</ObjectProperty>
```

```
<!-- http://www.semanticweb.org/ontologies/2012/11/Legal
Ontology1.owl#Has 伤情等级 -->
```

```
<ObjectProperty rdf:about="&LegalOntology1;Has 伤情等级">
  <rdfs:range rdf:resource="&LegalOntology1;伤情等级" />
  <rdfs:subPropertyOf rdf:resource="&LegalOntology1;案情类中的对象属性" />
  <rdfs:domain rdf:resource="&LegalOntology1;结果" />
</ObjectProperty>

<!-- http://www.semanticweb.org/ontologies/2012/11/LegalOntology1.owl#Has 作案工具 -->

<ObjectProperty rdf:about="&LegalOntology1;Has 作案工具">
  <rdfs:range rdf:resource="&LegalOntology1;作案工具" />
  <rdfs:subPropertyOf rdf:resource="&LegalOntology1;案情类中的对象属性" />
  <rdfs:domain rdf:resource="&LegalOntology1;经过" />
</ObjectProperty>

<!-- http://www.semanticweb.org/ontologies/2012/11/LegalOntology1.owl#Has 依据法条 -->

<ObjectProperty rdf:about="&LegalOntology1;Has 依据法条">
  <rdfs:domain rdf:resource="&LegalOntology1;判决结果" />
  <rdfs:subPropertyOf rdf:resource="&LegalOntology1;判决结果类中的对象属性" />
  <rdfs:range rdf:resource="&LegalOntology1;法条" />
</ObjectProperty>

<!-- http://www.semanticweb.org/ontologies/2012/11/LegalOntology1.owl#Has 公诉机关 -->

<ObjectProperty rdf:about="&LegalOntology1;Has 公诉机关">
```



```

        <rdfs:range rdf:resource="&LegalOntology1;检察院"/>
        <rdfs:domain rdf:resource="&LegalOntology1;起诉书"/>
        <rdfs:subPropertyOf rdf:resource="&LegalOntology1;起
起诉书类中的对象属性"/>
    </ObjectProperty>

    <!-- http://www.semanticweb.org/ontologies/2012/11/Legal
Ontology1.owl#Has 判决结果 -->

    <ObjectProperty rdf:about="&LegalOntology1;Has 判决结果">
        <rdfs:domain rdf:resource="&LegalOntology1;判决书"/>
        <rdfs:subPropertyOf rdf:resource="&LegalOntology1;判
判决书类中的对象属性"/>
        <rdfs:range rdf:resource="&LegalOntology1;判决结果"/>
    </ObjectProperty>

    <!-- http://www.semanticweb.org/ontologies/2012/11/Legal
Ontology1.owl#Has 合议庭 -->

    <ObjectProperty rdf:about="&LegalOntology1;Has 合议庭">
        <rdfs:domain rdf:resource="&LegalOntology1;判决书"/>
        <rdfs:subPropertyOf rdf:resource="&LegalOntology1;判
判决书类中的对象属性"/>
        <rdfs:range rdf:resource="&LegalOntology1;合议庭"/>
    </ObjectProperty>

    <!-- http://www.semanticweb.org/ontologies/2012/11/Legal
Ontology1.owl#Has 审判机构 -->

    <ObjectProperty rdf:about="&LegalOntology1;Has 审判机构">
        <rdfs:domain rdf:resource="&LegalOntology1;判决书"/>
        <rdfs:subPropertyOf rdf:resource="&LegalOntology1;判
判决书类中的对象属性"/>

```

```
<rdfs:range rdf:resource="&LegalOntology1;法院" />
</ObjectProperty>

<!-- http://www.semanticweb.org/ontologies/2012/11/Legal
Ontology1.owl#Has 攻击部位 -->

<ObjectProperty rdf:about="&LegalOntology1;Has 攻击部位">
  <rdfs:range rdf:resource="&LegalOntology1;攻击部位" />
  <rdfs:subPropertyOf rdf:resource="&LegalOntology1;案情类中的对象属性" />
  <rdfs:domain rdf:resource="&LegalOntology1;经过" />
</ObjectProperty>

<!-- http://www.semanticweb.org/ontologies/2012/11/Legal
Ontology1.owl#Has 案情 -->

<ObjectProperty rdf:about="&LegalOntology1;Has 案情">
  <rdfs:domain rdf:resource="&LegalOntology1;判决书" />
  <rdfs:subPropertyOf rdf:resource="&LegalOntology1;判决书类中的对象属性" />
  <rdfs:range rdf:resource="&LegalOntology1;案情" />
</ObjectProperty>

<!-- http://www.semanticweb.org/ontologies/2012/11/Legal
Ontology1.owl#Has 法定代理人 -->

<ObjectProperty rdf:about="&LegalOntology1;Has 法定代理人">
  <rdfs:range rdf:resource="&LegalOntology1;法定代理人" />
  <rdfs:domain rdf:resource="&LegalOntology1;被告人" />
  <rdfs:subPropertyOf rdf:resource="&LegalOntology1;被告人类中的对象属性" />
```

```

</ObjectProperty>

<!-- http://www.semanticweb.org/ontologies/2012/11/Legal
Ontology1.owl#Has 组织法院 -->

<ObjectProperty rdf:about="&LegalOntology1;Has 组织法院">
  <rdfs:domain rdf:resource="&LegalOntology1;合议庭"/>
  <rdfs:subPropertyOf rdf:resource="&LegalOntology1;合
合议庭类的对象属性"/>
  <rdfs:range rdf:resource="&LegalOntology1;法院"/>
</ObjectProperty>

<!-- http://www.semanticweb.org/ontologies/2012/11/Legal
Ontology1.owl#Has 被告人 -->

<ObjectProperty rdf:about="&LegalOntology1;Has 被告人">
  <rdfs:domain rdf:resource="&LegalOntology1;判决书"/>
  <rdfs:subPropertyOf rdf:resource="&LegalOntology1;判
判决书类中的对象属性"/>
  <rdfs:range rdf:resource="&LegalOntology1;被告人"/>
</ObjectProperty>

<!-- http://www.semanticweb.org/ontologies/2012/11/Legal
Ontology1.owl#Has 证据 -->

<ObjectProperty rdf:about="&LegalOntology1;Has 证据">
  <rdfs:domain rdf:resource="&LegalOntology1;判决书"/>
  <rdfs:subPropertyOf rdf:resource="&LegalOntology1;判
判决书类中的对象属性"/>
  <rdfs:range rdf:resource="&LegalOntology1;证据"/>
</ObjectProperty>

```

```
<!-- http://www.semanticweb.org/ontologies/2012/11/Legal
Ontology1.owl#Has 起诉信息 -->
```

```
<ObjectProperty rdf:about="&LegalOntology1;Has 起诉信息">
  <rdfs:domain rdf:resource="&LegalOntology1;判决书"/>
  <rdfs:subPropertyOf rdf:resource="&LegalOntology1;判
判决书类中的对象属性"/>
  <rdfs:range rdf:resource="&LegalOntology1;起诉书"/>
</ObjectProperty>
```

```
<!-- http://www.semanticweb.org/ontologies/2012/11/Legal
Ontology1.owl#Has 起诉信息_案情 -->
```

```
<ObjectProperty rdf:about="&LegalOntology1;Has 起诉信息_案情">
  <rdfs:range rdf:resource="&LegalOntology1;案情"/>
  <rdfs:domain rdf:resource="&LegalOntology1;起诉书"/>
  <rdfs:subPropertyOf rdf:resource="&LegalOntology1;起
诉书类中的对象属性"/>
</ObjectProperty>
```

```
<!-- http://www.semanticweb.org/ontologies/2012/11/Legal
Ontology1.owl#Has 起诉信息_被告人 -->
```

```
<ObjectProperty rdf:about="&LegalOntology1;Has 起诉信息_被
告人">
  <rdfs:domain rdf:resource="&LegalOntology1;起诉书"/>
  <rdfs:subPropertyOf rdf:resource="&LegalOntology1;起
诉书类中的对象属性"/>
  <rdfs:range rdf:resource="&LegalOntology1;起诉信息_被
告人"/>
</ObjectProperty>
```

```
<!-- http://www.semanticweb.org/ontologies/2012/11/Legal
```

Ontology1.owl#Has 起诉信息_证据 -->

```
<ObjectProperty rdf:about="&LegalOntology1;Has 起诉信息_证据">
  <rdfs:range rdf:resource="&LegalOntology1;证据"/>
  <rdfs:domain rdf:resource="&LegalOntology1;起诉书"/>
  <rdfs:subPropertyOf rdf:resource="&LegalOntology1;起
起诉书类中的对象属性"/>
</ObjectProperty>
```

<!-- http://www.semanticweb.org/ontologies/2012/11/Legal
Ontology1.owl#Has 辩护人 -->

```
<ObjectProperty rdf:about="&LegalOntology1;Has 辩护人">
  <rdfs:domain rdf:resource="&LegalOntology1;被告人"/>
  <rdfs:subPropertyOf rdf:resource="&LegalOntology1;被
告人类中的对象属性"/>
  <rdfs:range rdf:resource="&LegalOntology1;辩护人"/>
</ObjectProperty>
```

<!-- http://www.semanticweb.org/ontologies/2012/11/Legal
Ontology1.owl#判决书类中的对象属性 -->

```
<ObjectProperty rdf:about="&LegalOntology1;判决书类中的
对象属性"/>
```

<!-- http://www.semanticweb.org/ontologies/2012/11/Legal
Ontology1.owl#判决结果类中的对象属性 -->

```
<ObjectProperty rdf:about="&LegalOntology1;判决结果类中的
对象属性"/>
```

```
<!-- http://www.semanticweb.org/ontologies/2012/11/Legal
Ontology1.owl#合议庭类的对象属性 -->
```

```
<ObjectProperty rdf:about="&LegalOntology1;合议庭类的对象
属性"/>
```

```
<!-- http://www.semanticweb.org/ontologies/2012/11/Legal
Ontology1.owl#案情类中的对象属性 -->
```

```
<ObjectProperty rdf:about="&LegalOntology1;案情类中的对象
属性">
```

```
<rdfs:range rdf:resource="&LegalOntology1;从重情节"/>
```

```
<rdfs:domain rdf:resource="&LegalOntology1;结果"/>
```

```
</ObjectProperty>
```

```
<!-- http://www.semanticweb.org/ontologies/2012/11/Legal
Ontology1.owl#被告人类中的对象属性 -->
```

```
<ObjectProperty rdf:about="&LegalOntology1;被告人类中的对
象属性"/>
```

```
<!-- http://www.semanticweb.org/ontologies/2012/11/Legal
Ontology1.owl#起诉书类中的对象属性 -->
```

```
<ObjectProperty rdf:about="&LegalOntology1;起诉书类中的对
象属性"/>
```

```

<!--
////////////////////////////////////
//
// Data properties
//
////////////////////////////////////
-->

<!-- http://www.semanticweb.org/ontologies/2012/11/Legal
Ontology1.owl#与被告关系 -->

<DatatypeProperty rdf:about="&LegalOntology1;与被告关系">
    <rdfs:subPropertyOf rdf:resource="&LegalOntology1;法
定代理人子类数据属性"/>
</DatatypeProperty>

<!-- http://www.semanticweb.org/ontologies/2012/11/Legal
Ontology1.owl#书记员 -->

<DatatypeProperty rdf:about="&LegalOntology1;书记员">
    <rdfs:domain rdf:resource="&LegalOntology1;合议庭"/>
    <rdfs:subPropertyOf rdf:resource="&LegalOntology1;合
议庭类数据属性"/>
</DatatypeProperty>

<!-- http://www.semanticweb.org/ontologies/2012/11/Legal
Ontology1.owl#事件 -->

<DatatypeProperty rdf:about="&LegalOntology1;事件">
    <rdfs:subPropertyOf rdf:resource="&LegalOntology1;起
因子类数据属性"/>
</DatatypeProperty>

```

```
<!-- http://www.semanticweb.org/ontologies/2012/11/Legal
Ontology1.owl#人民陪审员 -->
```

```
<DatatypeProperty rdf:about="&LegalOntology1;人民陪审员">
  <rdfs:domain rdf:resource="&LegalOntology1;合议庭"/>
  <rdfs:subPropertyOf rdf:resource="&LegalOntology1;合
合议庭类数据属性"/>
</DatatypeProperty>
```

```
<!-- http://www.semanticweb.org/ontologies/2012/11/Legal
Ontology1.owl#人证子类数据属性 -->
```

```
<DatatypeProperty rdf:about="&LegalOntology1;人证子类数据
属性">
  <rdfs:subPropertyOf rdf:resource="&LegalOntology1;证
据类数据属性"/>
</DatatypeProperty>
```

```
<!-- http://www.semanticweb.org/ontologies/2012/11/Legal
Ontology1.owl#代理审判员 -->
```

```
<DatatypeProperty rdf:about="&LegalOntology1;代理审判员">
  <rdfs:domain rdf:resource="&LegalOntology1;合议庭"/>
  <rdfs:subPropertyOf rdf:resource="&LegalOntology1;合
合议庭类数据属性"/>
</DatatypeProperty>
```

```
<!-- http://www.semanticweb.org/ontologies/2012/11/Legal
Ontology1.owl#代理审判员长 -->
```

```
<DatatypeProperty rdf:about="&LegalOntology1;代理审判员长">
  <rdfs:domain rdf:resource="&LegalOntology1;合议庭"/>
```



```

        <rdfs:subPropertyOf rdf:resource="&LegalOntology1;合
议庭类数据属性"/>
    </DatatypeProperty>

    <!-- http://www.semanticweb.org/ontologies/2012/11/Legal
Ontology1.owl#先施暴方 -->

    <DatatypeProperty rdf:about="&LegalOntology1;先施暴方">
        <rdfs:subPropertyOf rdf:resource="&LegalOntology1;起
因子类数据属性"/>
    </DatatypeProperty>

    <!-- http://www.semanticweb.org/ontologies/2012/11/Legal
Ontology1.owl#其他附加 -->

    <DatatypeProperty rdf:about="&LegalOntology1;其他附加">
        <rdfs:subPropertyOf rdf:resource="&LegalOntology1;判
决结果类数据属性"/>
    </DatatypeProperty>

    <!-- http://www.semanticweb.org/ontologies/2012/11/Legal
Ontology1.owl#刑期 -->

    <DatatypeProperty rdf:about="&LegalOntology1;刑期">
        <rdfs:subPropertyOf rdf:resource="&LegalOntology1;判
决结果类数据属性"/>
    </DatatypeProperty>

    <!-- http://www.semanticweb.org/ontologies/2012/11/Legal
Ontology1.owl#判决书类数据属性 -->

    <DatatypeProperty rdf:about="&LegalOntology1;判决书类数据
属性"/>

```

```
<!-- http://www.semanticweb.org/ontologies/2012/11/Legal
Ontology1.owl#判决书编号 -->
```

```
<DatatypeProperty rdf:about="&LegalOntology1;判决书编号">
    <rdfs:subPropertyOf rdf:resource="&LegalOntology1;判
    决书类数据属性"/>
```

```
</DatatypeProperty>
```

```
<!-- http://www.semanticweb.org/ontologies/2012/11/Legal
Ontology1.owl#判决结果类数据属性 -->
```

```
<DatatypeProperty rdf:about="&LegalOntology1;判决结果类数
据属性"/>
```

```
<!-- http://www.semanticweb.org/ontologies/2012/11/Legal
Ontology1.owl#受害方 -->
```

```
<DatatypeProperty rdf:about="&LegalOntology1;受害方">
    <rdfs:subPropertyOf rdf:resource="&LegalOntology1;经
    过子类数据属性"/>
```

```
</DatatypeProperty>
```

```
<!-- http://www.semanticweb.org/ontologies/2012/11/Legal
Ontology1.owl#受害方人数 -->
```

```
<DatatypeProperty rdf:about="&LegalOntology1;受害方人数">
    <rdfs:subPropertyOf rdf:resource="&LegalOntology1;经
    过子类数据属性"/>
```

```
</DatatypeProperty>
```

```
<!-- http://www.semanticweb.org/ontologies/2012/11/Legal
Ontology1.owl#合议庭类数据属性 -->
```

```
<DatatypeProperty rdf:about="&LegalOntology1;合议庭类数据
属性"/>
```

```
<!-- http://www.semanticweb.org/ontologies/2012/11/Legal
Ontology1.owl#审判员 -->
```

```
<DatatypeProperty rdf:about="&LegalOntology1;审判员">
  <rdfs:domain rdf:resource="&LegalOntology1;合议庭"/>
  <rdfs:subPropertyOf rdf:resource="&LegalOntology1;合
合议庭类数据属性"/>
</DatatypeProperty>
```

```
<!-- http://www.semanticweb.org/ontologies/2012/11/Legal
Ontology1.owl#审判长 -->
```

```
<DatatypeProperty rdf:about="&LegalOntology1;审判长">
  <rdfs:domain rdf:resource="&LegalOntology1;合议庭"/>
  <rdfs:subPropertyOf rdf:resource="&LegalOntology1;合
合议庭类数据属性"/>
</DatatypeProperty>
```

```
<!-- http://www.semanticweb.org/ontologies/2012/11/Legal
Ontology1.owl#审理状态 -->
```

```
<DatatypeProperty rdf:about="&LegalOntology1;审理状态">
  <rdfs:subPropertyOf rdf:resource="&LegalOntology1;判
决书类数据属性"/>
</DatatypeProperty>
```

```
<!-- http://www.semanticweb.org/ontologies/2012/11/Legal
Ontology1.owl#归属 -->
```

```
<DatatypeProperty rdf:about="&LegalOntology1;归属">
    <rdfs:subPropertyOf rdf:resource="&LegalOntology1;物
证子类数据属性"/>
</DatatypeProperty>
```

```
<!-- http://www.semanticweb.org/ontologies/2012/11/Legal
Ontology1.owl#提起公诉日期 -->
```

```
<DatatypeProperty rdf:about="&LegalOntology1;提起公诉日期">
    <rdfs:subPropertyOf rdf:resource="&LegalOntology1;起
诉书类数据属性"/>
</DatatypeProperty>
```

```
<!-- http://www.semanticweb.org/ontologies/2012/11/Legal
Ontology1.owl#攻击次数 -->
```

```
<DatatypeProperty rdf:about="&LegalOntology1;攻击次数">
    <rdfs:subPropertyOf rdf:resource="&LegalOntology1;经
过子类数据属性"/>
</DatatypeProperty>
```

```
<!-- http://www.semanticweb.org/ontologies/2012/11/Legal
Ontology1.owl#施暴方式 -->
```

```
<DatatypeProperty rdf:about="&LegalOntology1;施暴方式">
    <rdfs:subPropertyOf rdf:resource="&LegalOntology1;经
过子类数据属性"/>
</DatatypeProperty>
```

```
<!-- http://www.semanticweb.org/ontologies/2012/11/Legal
```

Ontology1.owl#案情_被告人姓名 -->

```
<DatatypeProperty rdf:about="&LegalOntology1;案情_被告人姓名">
  <rdfs:subPropertyOf rdf:resource="&LegalOntology1;案情类数据属性"/>
</DatatypeProperty>
```

<!-- http://www.semanticweb.org/ontologies/2012/11/Legal
Ontology1.owl#案情类数据属性 -->

```
<DatatypeProperty rdf:about="&LegalOntology1;案情类数据属性"/>
```

<!-- http://www.semanticweb.org/ontologies/2012/11/Legal
Ontology1.owl#检察员 -->

```
<DatatypeProperty rdf:about="&LegalOntology1;检察员">
  <rdfs:domain rdf:resource="&LegalOntology1;合议庭"/>
  <rdfs:subPropertyOf rdf:resource="&LegalOntology1;合议庭类数据属性"/>
</DatatypeProperty>
```

<!-- http://www.semanticweb.org/ontologies/2012/11/Legal
Ontology1.owl#法定代理人住址 -->

```
<DatatypeProperty rdf:about="&LegalOntology1;法定代理人住址">
  <rdfs:subPropertyOf rdf:resource="&LegalOntology1;法定代理人子类数据属性"/>
</DatatypeProperty>
```

```
<!-- http://www.semanticweb.org/ontologies/2012/11/Legal
Ontology1.owl#法定代理人出生日期 -->
```

```
<DatatypeProperty rdf:about="&LegalOntology1;法定代理人出
生日期">
```

```
<rdfs:subPropertyOf rdf:resource="&LegalOntology1;法
定代理人子类数据属性"/>
```

```
</DatatypeProperty>
```

```
<!-- http://www.semanticweb.org/ontologies/2012/11/Legal
Ontology1.owl#法定代理人姓名 -->
```

```
<DatatypeProperty rdf:about="&LegalOntology1;法定代理人姓名">
```

```
<rdfs:subPropertyOf rdf:resource="&LegalOntology1;法
定代理人子类数据属性"/>
```

```
</DatatypeProperty>
```

```
<!-- http://www.semanticweb.org/ontologies/2012/11/Legal
Ontology1.owl#法定代理人子类数据属性 -->
```

```
<DatatypeProperty rdf:about="&LegalOntology1;法定代理人子
类数据属性">
```

```
<rdfs:subPropertyOf rdf:resource="&LegalOntology1;法
律角色类数据属性"/>
```

```
</DatatypeProperty>
```

```
<!-- http://www.semanticweb.org/ontologies/2012/11/Legal
Ontology1.owl#法定代理人年龄 -->
```

```
<DatatypeProperty rdf:about="&LegalOntology1;法定代理人年龄">
```

```
<rdfs:subPropertyOf rdf:resource="&LegalOntology1;法
定代理人子类数据属性"/>
```

```

</DatatypeProperty>

<!-- http://www.semanticweb.org/ontologies/2012/11/Legal
Ontology1.owl#法定代理人民族 -->

<DatatypeProperty rdf:about="&LegalOntology1;法定代理人民族">
  <rdfs:subPropertyOf rdf:resource="&LegalOntology1;法
定代理人子类数据属性"/>
</DatatypeProperty>

<!-- http://www.semanticweb.org/ontologies/2012/11/Legal
Ontology1.owl#法定代理人职业 -->

<DatatypeProperty rdf:about="&LegalOntology1;法定代理人职业">
  <rdfs:subPropertyOf rdf:resource="&LegalOntology1;法
定代理人子类数据属性"/>
</DatatypeProperty>

<!-- http://www.semanticweb.org/ontologies/2012/11/Legal
Ontology1.owl#法律角色类数据属性 -->

<DatatypeProperty rdf:about="&LegalOntology1;法律角色类数
据属性"/>

<!-- http://www.semanticweb.org/ontologies/2012/11/Legal
Ontology1.owl#法条内容 -->

<DatatypeProperty rdf:about="&LegalOntology1;法条内容">
  <rdfs:subPropertyOf rdf:resource="&LegalOntology1;法
条类数据属性"/>
</DatatypeProperty>

<!-- http://www.semanticweb.org/ontologies/2012/11/Legal

```

Ontology1.owl#法条类数据属性 -->

```
<DatatypeProperty rdf:about="&LegalOntology1;法条类数据属性"/>
```

<!-- http://www.semanticweb.org/ontologies/2012/11/Legal
Ontology1.owl#法条编号 -->

```
<DatatypeProperty rdf:about="&LegalOntology1;法条编号">  
  <rdfs:subPropertyOf rdf:resource="&LegalOntology1;法  
条类数据属性"/>  
</DatatypeProperty>
```

<!-- http://www.semanticweb.org/ontologies/2012/11/Legal
Ontology1.owl#涉嫌罪名 -->

```
<DatatypeProperty rdf:about="&LegalOntology1;涉嫌罪名">  
  <rdfs:subPropertyOf rdf:resource="&LegalOntology1;被  
告人子类数据属性"/>  
</DatatypeProperty>
```

<!-- http://www.semanticweb.org/ontologies/2012/11/Legal
Ontology1.owl#物品 -->

```
<DatatypeProperty rdf:about="&LegalOntology1;物品">  
  <rdfs:subPropertyOf rdf:resource="&LegalOntology1;物  
证子类数据属性"/>  
</DatatypeProperty>
```

<!-- http://www.semanticweb.org/ontologies/2012/11/Legal
Ontology1.owl#物证子类数据属性 -->

```
<DatatypeProperty rdf:about="&LegalOntology1;物证子类数据  
属性">
```



```

        <rdfs:subPropertyOf rdf:resource="&LegalOntology1;证
据类数据属性"/>
    </DatatypeProperty>

    <!-- http://www.semanticweb.org/ontologies/2012/11/Legal
Ontology1.owl#犯案地点 -->

    <DatatypeProperty rdf:about="&LegalOntology1;犯案地点">
        <rdfs:subPropertyOf rdf:resource="&LegalOntology1;案
情类数据属性"/>
    </DatatypeProperty>

    <!-- http://www.semanticweb.org/ontologies/2012/11/Legal
Ontology1.owl#犯案时间 -->

    <DatatypeProperty rdf:about="&LegalOntology1;犯案时间">
        <rdfs:subPropertyOf rdf:resource="&LegalOntology1;案
情类数据属性"/>
    </DatatypeProperty>

    <!-- http://www.semanticweb.org/ontologies/2012/11/Legal
Ontology1.owl#现羁押地 -->

    <DatatypeProperty rdf:about="&LegalOntology1;现羁押地">
        <rdfs:subPropertyOf rdf:resource="&LegalOntology1;被
告人子类数据属性"/>
    </DatatypeProperty>

    <!-- http://www.semanticweb.org/ontologies/2012/11/Legal
Ontology1.owl#经过子类数据属性 -->

```

```
<DatatypeProperty rdf:about="&LegalOntology1;经过子类数据  
属性">
```

```
<rdfs:subPropertyOf rdf:resource="&LegalOntology1;案情  
类数据属性"/>
```

```
</DatatypeProperty>
```

```
<!-- http://www.semanticweb.org/ontologies/2012/11/Legal  
Ontology1.owl#罪名 -->
```

```
<DatatypeProperty rdf:about="&LegalOntology1;罪名">
```

```
<rdfs:subPropertyOf rdf:resource="&LegalOntology1;判  
决结果类数据属性"/>
```

```
</DatatypeProperty>
```

```
<!-- http://www.semanticweb.org/ontologies/2012/11/Legal  
Ontology1.owl#获取物证时间 -->
```

```
<DatatypeProperty rdf:about="&LegalOntology1;获取物证时间">
```

```
<rdfs:subPropertyOf rdf:resource="&LegalOntology1;物  
证子类数据属性"/>
```

```
</DatatypeProperty>
```

```
<!-- http://www.semanticweb.org/ontologies/2012/11/Legal  
Ontology1.owl#被告人住址 -->
```

```
<DatatypeProperty rdf:about="&LegalOntology1;被告人住址">
```

```
<rdfs:subPropertyOf rdf:resource="&LegalOntology1;被  
告人子类数据属性"/>
```

```
</DatatypeProperty>
```

```
<!-- http://www.semanticweb.org/ontologies/2012/11/Legal
Ontology1.owl#被告人出生年月 -->
```

```
<DatatypeProperty rdf:about="&LegalOntology1;被告人出生年月">
    <rdfs:subPropertyOf rdf:resource="&LegalOntology1;被
告人子类数据属性"/>
</DatatypeProperty>
```

```
<!-- http://www.semanticweb.org/ontologies/2012/11/Legal
Ontology1.owl#被告人姓名 -->
```

```
<DatatypeProperty rdf:about="&LegalOntology1;被告人姓名">
    <rdfs:subPropertyOf rdf:resource="&LegalOntology1;被
告人子类数据属性"/>
</DatatypeProperty>
```

```
<!-- http://www.semanticweb.org/ontologies/2012/11/Legal
Ontology1.owl#被告人子类数据属性 -->
```

```
<DatatypeProperty rdf:about="&LegalOntology1;被告人子类数
据属性">
    <rdfs:subPropertyOf rdf:resource="&LegalOntology1;法
律角色类数据属性"/>
</DatatypeProperty>
```

```
<!-- http://www.semanticweb.org/ontologies/2012/11/Legal
Ontology1.owl#被告人年龄 -->
```

```
<DatatypeProperty rdf:about="&LegalOntology1;被告人年龄">  
    <rdfs:subPropertyOf rdf:resource="&LegalOntology1;被告  
被告人子类数据属性"/>  
</DatatypeProperty>
```

```
<!-- http://www.semanticweb.org/ontologies/2012/11/Legal  
Ontology1.owl#被告人户籍地 -->
```

```
<DatatypeProperty rdf:about="&LegalOntology1;被告人户籍地">  
    <rdfs:subPropertyOf rdf:resource="&LegalOntology1;被告  
被告人子类数据属性"/>  
</DatatypeProperty>
```

```
<!-- http://www.semanticweb.org/ontologies/2012/11/Legal  
Ontology1.owl#被告人文化水平 -->
```

```
<DatatypeProperty rdf:about="&LegalOntology1;被告人文化水平">  
    <rdfs:subPropertyOf rdf:resource="&LegalOntology1;被告  
被告人子类数据属性"/>  
</DatatypeProperty>
```

```
<!-- http://www.semanticweb.org/ontologies/2012/11/Legal  
Ontology1.owl#被告人民族 -->
```

```
<DatatypeProperty rdf:about="&LegalOntology1;被告人民族">  
    <rdfs:subPropertyOf rdf:resource="&LegalOntology1;被告  
被告人子类数据属性"/>  
</DatatypeProperty>
```

```
<!-- http://www.semanticweb.org/ontologies/2012/11/Legal
Ontology1.owl#被告人职业 -->
```

```
<DatatypeProperty rdf:about="&LegalOntology1;被告人职业">
  <rdfs:subPropertyOf rdf:resource="&LegalOntology1;被
告人子类数据属性"/>
</DatatypeProperty>
```

```
<!-- http://www.semanticweb.org/ontologies/2012/11/Legal
Ontology1.owl#被害人 -->
```

```
<DatatypeProperty rdf:about="&LegalOntology1;被害人">
  <rdfs:subPropertyOf rdf:resource="&LegalOntology1;案
情类数据属性"/>
</DatatypeProperty>
```

```
<!-- http://www.semanticweb.org/ontologies/2012/11/Legal
Ontology1.owl#被羁押日期 -->
```

```
<DatatypeProperty rdf:about="&LegalOntology1;被羁押日期">
  <rdfs:subPropertyOf rdf:resource="&LegalOntology1;被
告人子类数据属性"/>
</DatatypeProperty>
```

```
<!-- http://www.semanticweb.org/ontologies/2012/11/Legal
Ontology1.owl#被逮捕日期 -->
```

```
<DatatypeProperty rdf:about="&LegalOntology1;被逮捕日期">
    <rdfs:subPropertyOf rdf:resource="&LegalOntology1;被
告人子类数据属性"/>
</DatatypeProperty>
```

```
<!-- http://www.semanticweb.org/ontologies/2012/11/Legal
Ontology1.owl#证人涉案关系 -->
```

```
<DatatypeProperty rdf:about="&LegalOntology1;证人涉案关系">
    <rdfs:subPropertyOf rdf:resource="&LegalOntology1;人
证子类数据属性"/>
</DatatypeProperty>
```

```
<!-- http://www.semanticweb.org/ontologies/2012/11/Legal
Ontology1.owl#证人证明内容 -->
```

```
<DatatypeProperty rdf:about="&LegalOntology1;证人证明内容">
    <rdfs:subPropertyOf rdf:resource="&LegalOntology1;人
证子类数据属性"/>
</DatatypeProperty>
```

```
<!-- http://www.semanticweb.org/ontologies/2012/11/Legal
Ontology1.owl#证人证明地点 -->
```

```
<DatatypeProperty rdf:about="&LegalOntology1;证人证明地点">
    <rdfs:subPropertyOf rdf:resource="&LegalOntology1;人
证子类数据属性"/>
</DatatypeProperty>
```

```
<!-- http://www.semanticweb.org/ontologies/2012/11/Legal
Ontology1.owl#证人证明时间 -->
```

```
<DatatypeProperty rdf:about="&LegalOntology1;证人证明时间">
  <rdfs:subPropertyOf rdf:resource="&LegalOntology1;人
证子类数据属性"/>
</DatatypeProperty>
```

```
<!-- http://www.semanticweb.org/ontologies/2012/11/Legal
Ontology1.owl#证据类数据属性 -->
```

```
<DatatypeProperty rdf:about="&LegalOntology1;证据类数据
属性"/>
```

```
<!-- http://www.semanticweb.org/ontologies/2012/11/Legal
Ontology1.owl#起因子类数据属性 -->
```

```
<DatatypeProperty rdf:about="&LegalOntology1;起因子类数据
属性">
  <rdfs:subPropertyOf rdf:resource="&LegalOntology1;案
情类数据属性"/>
</DatatypeProperty>
```

```
<!-- http://www.semanticweb.org/ontologies/2012/11/Legal
Ontology1.owl#起诉书类数据属性 -->
```

```
<DatatypeProperty rdf:about="&LegalOntology1;起诉书类数据
属性"/>
```

```
<!-- http://www.semanticweb.org/ontologies/2012/11/Legal
Ontology1.owl#起诉书编号 -->
```

```
<DatatypeProperty rdf:about="&LegalOntology1;起诉书编号">
    <rdfs:subPropertyOf rdf:resource="&LegalOntology1;起
    诉书类数据属性"/>
</DatatypeProperty>
```

```
<!-- http://www.semanticweb.org/ontologies/2012/11/Legal
Ontology1.owl#起诉信息_指控罪名 -->
```

```
<DatatypeProperty rdf:about="&LegalOntology1;起诉信息_指控
罪名">
    <rdfs:subPropertyOf rdf:resource="&LegalOntology1;起
    诉信息_被告人子类数据属性"/>
</DatatypeProperty>
```

```
<!-- http://www.semanticweb.org/ontologies/2012/11/Legal
Ontology1.owl#起诉信息_被告人姓名 -->
```

```
<DatatypeProperty rdf:about="&LegalOntology1;起诉信息_被告
人姓名">
    <rdfs:subPropertyOf rdf:resource="&LegalOntology1;起
    诉信息_被告人子类数据属性"/>
</DatatypeProperty>
```

```
<!-- http://www.semanticweb.org/ontologies/2012/11/Legal
Ontology1.owl#起诉信息_被告人子类数据属性 -->
```



```
<DatatypeProperty rdf:about="&LegalOntology1;起诉信息_被告人子类数据属性">
```

```
<rdfs:subPropertyOf rdf:resource="&LegalOntology1;被告人子类数据属性"/>
```

```
</DatatypeProperty>
```

```
<!-- http://www.semanticweb.org/ontologies/2012/11/LegalOntology1.owl#辩护人单位 -->
```

```
<DatatypeProperty rdf:about="&LegalOntology1;辩护人单位">
```

```
<rdfs:subPropertyOf rdf:resource="&LegalOntology1;辩护人子类数据属性"/>
```

```
</DatatypeProperty>
```

```
<!-- http://www.semanticweb.org/ontologies/2012/11/LegalOntology1.owl#辩护人姓名 -->
```

```
<DatatypeProperty rdf:about="&LegalOntology1;辩护人姓名">
```

```
<rdfs:subPropertyOf rdf:resource="&LegalOntology1;辩护人子类数据属性"/>
```

```
</DatatypeProperty>
```

```
<!-- http://www.semanticweb.org/ontologies/2012/11/LegalOntology1.owl#辩护人子类数据属性 -->
```

```
<DatatypeProperty rdf:about="&LegalOntology1;辩护人子类数据属性">
```

```
<rdfs:subPropertyOf rdf:resource="&LegalOntology1;法
```

律角色类数据属性"/>

```
</DatatypeProperty>
```

```
<!--
```

```
////////////////////////////////////
```

```
//
```

```
// Classes
```

```
//
```

```
////////////////////////////////////
```

```
-->
```

```
<!-- http://www.semanticweb.org/ontologies/2012/11/Legal
Ontology1.owl#人证 -->
```

```
<Class rdf:about="&LegalOntology1;人证">
```

```
<rdfs:subClassOf rdf:resource="&LegalOntology1;证据"/>
```

```
</Class>
```

```
<!-- http://www.semanticweb.org/ontologies/2012/11/Legal
Ontology1.owl#从轻情节 -->
```

```
<Class rdf:about="&LegalOntology1;从轻情节">
```

```
<rdfs:subClassOf rdf:resource="&LegalOntology1;案情相
关类"/>
```

```
</Class>
```

```
<!-- http://www.semanticweb.org/ontologies/2012/11/Legal
Ontology1.owl#从重情节 -->
```

```

<Class rdf:about="&LegalOntology1;从重情节">
    <rdfs:subClassOf rdf:resource="&LegalOntology1;案情相
    关类"/>
</Class>

```

```

<!-- http://www.semanticweb.org/ontologies/2012/11/Legal
Ontology1.owl#伤情等级 -->

```

```

<Class rdf:about="&LegalOntology1;伤情等级">
    <rdfs:subClassOf rdf:resource="&LegalOntology1;案情相
    关类"/>
</Class>

```

```

<!-- http://www.semanticweb.org/ontologies/2012/11/Legal
Ontology1.owl#作案工具 -->

```

```

<Class rdf:about="&LegalOntology1;作案工具">
    <rdfs:subClassOf rdf:resource="&LegalOntology1;案情相
    关类"/>
</Class>

```

```

<!-- http://www.semanticweb.org/ontologies/2012/11/Legal
Ontology1.owl#判决书 -->

```

```

<Class rdf:about="&LegalOntology1;判决书">
    <rdfs:subClassOf rdf:resource="&owl;Thing"/>
</Class>

```

```
<!-- http://www.semanticweb.org/ontologies/2012/11/Legal
Ontology1.owl#判决结果 -->
```

```
<Class rdf:about="&LegalOntology1;判决结果">
    <rdfs:subClassOf rdf:resource="&owl;Thing"/>
</Class>
```

```
<!-- http://www.semanticweb.org/ontologies/2012/11/Legal
Ontology1.owl#合议庭 -->
```

```
<Class rdf:about="&LegalOntology1;合议庭">
    <rdfs:subClassOf rdf:resource="&owl;Thing"/>
</Class>
```

```
<!-- http://www.semanticweb.org/ontologies/2012/11/Legal
Ontology1.owl#攻击部位 -->
```

```
<Class rdf:about="&LegalOntology1;攻击部位">
    <rdfs:subClassOf rdf:resource="&LegalOntology1;案情相
关类"/>
</Class>
```

```
<!-- http://www.semanticweb.org/ontologies/2012/11/Legal
Ontology1.owl#案情 -->
```

```
<Class rdf:about="&LegalOntology1;案情">
    <rdfs:subClassOf rdf:resource="&owl;Thing"/>
```

```
</Class>
```

```
<!-- http://www.semanticweb.org/ontologies/2012/11/Legal  
Ontology1.owl#案情相关类 -->
```

```
<Class rdf:about="&LegalOntology1;案情相关类">  
  <rdfs:subClassOf rdf:resource="&LegalOntology1;案情"/>  
</Class>
```

```
<!-- http://www.semanticweb.org/ontologies/2012/11/Legal  
Ontology1.owl#检察院 -->
```

```
<Class rdf:about="&LegalOntology1;检察院">  
  <rdfs:subClassOf rdf:resource="&LegalOntology1;法律  
机构"/>  
</Class>
```

```
<!-- http://www.semanticweb.org/ontologies/2012/11/Legal  
Ontology1.owl#法定代理人 -->
```

```
<Class rdf:about="&LegalOntology1;法定代理人">  
  <rdfs:subClassOf rdf:resource="&LegalOntology1;法律  
角色"/>  
</Class>
```

```
<!-- http://www.semanticweb.org/ontologies/2012/11/Legal  
Ontology1.owl#法律机构 -->
```

```
<Class rdf:about="&LegalOntology1;法律机构"/>
```

```
<!-- http://www.semanticweb.org/ontologies/2012/11/Legal  
Ontology1.owl#法律角色 -->
```

```
<Class rdf:about="&LegalOntology1;法律角色">  
  <rdfs:subClassOf rdf:resource="&owl;Thing"/>  
</Class>
```

```
<!-- http://www.semanticweb.org/ontologies/2012/11/Legal  
Ontology1.owl#法条 -->
```

```
<Class rdf:about="&LegalOntology1;法条"/>
```

```
<!-- http://www.semanticweb.org/ontologies/2012/11/Legal  
Ontology1.owl#法院 -->
```

```
<Class rdf:about="&LegalOntology1;法院">  
  <rdfs:subClassOf rdf:resource="&LegalOntology1;法律  
机构"/>  
</Class>
```

```
<!-- http://www.semanticweb.org/ontologies/2012/11/Legal  
Ontology1.owl#物证 -->
```

```
<Class rdf:about="&LegalOntology1;物证">  
  <rdfs:subClassOf rdf:resource="&LegalOntology1;证据"/>
```

```
</Class>
```

```
<!-- http://www.semanticweb.org/ontologies/2012/11/Legal  
Ontology1.owl#经过 -->
```

```
<Class rdf:about="&LegalOntology1;经过">  
  <rdfs:subClassOf rdf:resource="&LegalOntology1;案情"/>  
</Class>
```

```
<!-- http://www.semanticweb.org/ontologies/2012/11/Legal  
Ontology1.owl#结果 -->
```

```
<Class rdf:about="&LegalOntology1;结果">  
  <rdfs:subClassOf rdf:resource="&LegalOntology1;案情"/>  
</Class>
```

```
<!-- http://www.semanticweb.org/ontologies/2012/11/Legal  
Ontology1.owl#被告人 -->
```

```
<Class rdf:about="&LegalOntology1;被告人">  
  <rdfs:subClassOf rdf:resource="&LegalOntology1;法律  
角色"/>  
</Class>
```

```
<!-- http://www.semanticweb.org/ontologies/2012/11/Legal  
Ontology1.owl#证据 -->
```

```
<Class rdf:about="&LegalOntology1;证据">  
  <rdfs:subClassOf rdf:resource="&owl;Thing"/>
```

```
</Class>

<!-- http://www.semanticweb.org/ontologies/2012/11/Legal
Ontology1.owl#起因 -->

<Class rdf:about="&LegalOntology1;起因">
  <rdfs:subClassOf rdf:resource="&LegalOntology1;案情"/>
</Class>

<!-- http://www.semanticweb.org/ontologies/2012/11/Legal
Ontology1.owl#起诉书 -->

<Class rdf:about="&LegalOntology1;起诉书">
  <rdfs:subClassOf rdf:resource="&owl;Thing"/>
</Class>

<!-- http://www.semanticweb.org/ontologies/2012/11/Legal
Ontology1.owl#起诉信息_被告人 -->

<Class rdf:about="&LegalOntology1;起诉信息_被告人">
  <rdfs:subClassOf rdf:resource="&LegalOntology1;被告人"/>
</Class>

<!-- http://www.semanticweb.org/ontologies/2012/11/Legal
Ontology1.owl#辩护人 -->

<Class rdf:about="&LegalOntology1;辩护人">
  <rdfs:subClassOf rdf:resource="&LegalOntology1;法律
角色"/>
</Class>
</rdf:RDF>
```


故意伤害罪类刑事案件本体

$O_{\text{故意伤害罪类刑事案件}} = \langle T, X, XD, AA, TR \rangle$, 其中, $T = \langle TC, TD, TO \rangle$

$TC = \{ \text{法律机构, 法院, 检察院, 起诉书, 合议庭, 法律角色, 被告人, 起诉信息_被告人, 辩护人, 法定代理人, 案情, 起因, 经过, 结果, 案情相关类, 攻击部位, 作案工具, 伤情等级, 从轻情节, 从重情节, 证据, 人证, 物证, 判决结果, 法条, 判决书} \}$

$TD = \{ \text{起诉书编号, 提起公诉日期, 审判长, 审判员, 代理审判员, 人民陪审员, 书记员, 检察员, 被告人姓名, 被告人性别, 被告人年龄, 被告人出生年月, 被告人民族, 被告人住址, 被告人户籍地, 被告人文化水平, 被告人职业, 涉嫌罪名, 被羁押日期, 被逮捕日期, 现羁押地, 是否累犯, 起诉信息_被告人姓名, 起诉信息_指控罪名, 辩护人姓名, 辩护人单位, 法定代理人姓名, 法定代理人年龄, 法定代理人出生日期, 法定代理人职业, 法定代理人住址, 与被告人关系, 案情_被告人姓名, 犯案地点, 被害人, 事件,} \}$

先施暴方, 施暴方式, 受害方, 受害人个数, 攻击次数, 获取物证时间, 归属, 物品, 证人证明时间, 证人涉案关系, 证人证明地点, 证人证明内容, 罪名, 刑期, 其他附加, 法条编号, 法条内容, 审理状态, 判决书编号 }

$TO = \{ \text{Has 起诉信息_被告人, Has 公诉机关, Has 起诉信息_案情, Has 起诉信息_证据, Has 法定代理人, Has 组织法院, Has 辩护人, Has 攻击部位, Has 作案工具, Has 伤情等级, Has 从轻情节, Has 从重情节, Has 依据法条, Has 起诉信息, Has 审判机构, Has 合议庭, Has 被告人, Has 案情, Has 证据, Has 判决结果} \}$

$X = \{ (\text{法院}_1, \text{法院}_2, \dots, \text{法院}_n), (\text{检察院}_1, \text{检察院}_2, \dots, \text{检察院}_n), (\text{起诉书}_1, \text{起诉书}_2, \dots, \text{起诉书}_n), (\text{合议庭}_1, \text{合议庭}_2, \dots, \text{合议庭}_n), (\text{被告人}_1, \text{被告人}_2, \dots, \text{被告人}_n), (\text{起诉信息_被告人}_1, \text{起诉信息_被告人}_2, \dots, \text{起诉信息_被告人}_n), (\text{辩护人}_1, \text{辩护人}_2, \dots, \text{辩护人}_n), (\text{法定代理人}_1, \text{法定代理人}_2, \dots, \text{法定代理人}_n), (\text{案情}_1, \text{案情}_2, \dots, \text{案情}_n), (\text{起因}_1, \text{起因}_2, \dots, \text{起因}_n), (\text{经过}_1, \text{经过}_2, \dots, \text{经过}_n), (\text{攻击部位}_1, \text{攻击部位}_2, \dots, \text{攻击部位}_n), (\text{作案工具}_1, \text{作案工具}_2, \dots, \text{作案工具}_n), (\text{证据}_1, \text{证据}_2, \dots, \text{证据}_n), (\text{人证}_1, \text{人证}_2, \dots, \text{人证}_n), (\text{物证}_1, \text{物证}_2, \dots, \text{物证}_n), (\text{判决结果}_1, \text{判决结果}_2, \dots, \text{判决结果}_n), (\text{法条}_1, \text{法条}_2, \dots, \text{法条}_n), (\text{判决书}_1, \text{判决书}_2, \dots, \text{判决书}_n) \}$

$XD = \{ \text{法院}(\text{法院}_i), \text{检察院}(\text{检察院}_i), \text{起诉书}(\text{起诉书}_i), \text{合议庭}(\text{合议庭}_i), \text{被告人}(\text{被告人}_i), \text{起诉信息_被告人}(\text{起诉信息_被告人}_i), \text{辩护人}(\text{辩护人}_i), \text{法定代理人}(\text{法定代理人}_i), \text{案情}(\text{案情}_i), \text{起因}(\text{起因}_i), \text{经过}(\text{经过}_i), \text{攻击部位}(\text{攻击部位}_i), \text{作案工具}(\text{作案工具}_i), \text{证据}(\text{证据}_i), \text{人证}(\text{人证}_i), \text{物证}(\text{物证}_i), \text{判决结果}(\text{判决结果}_i), \text{法条}(\text{法条}_i), \text{判决书}(\text{判决书}_i) \}$

AA = < AAO, AAC >

AAO = < Has 起诉信息_被告人(起诉书, 起诉信息_被告人), Has 公诉机关(起诉书, 检察院), Has 起诉信息_案情(起诉书, 案情), Has 起诉信息_证据(起诉书, 证据), Has 组织法院(合议庭, 法院), Has 法定代理人(被告人, 法定代理人), Has 辩护人(被告人, 辩护人), Has 攻击部位(经过, 攻击部位), Has 作案工具(经过, 作案工具), Has 伤情等级(结果, 伤情等级), Has 从轻情节(结果, 从轻情节), Has 从重情节(结果, 从重情节), Has 依据法条(判决结果, 法条), Has 起诉信息(判决书, 起诉书), Has 审判机构(判决书, 法院), Has 合议庭(判决书, 合议庭), Has 被告人(判决书, 被告人), Has 案情(判决书, 案情), Has 证据(判决书, 证据), Has 判决结果(判决书, 判决结果) >

AAC = < 起诉书 \sqsubseteq 起诉书编号.xsd:String \cap 提起公诉日期.xsd:String \cap Has 起诉信息_被告人 \cap Has 公诉机关 \cap Has 起诉信息_案情 \cap Has 起诉信息_证据 \cap , 合议庭 \sqsubseteq 审判长.xsd:String \cap 审判员.xsd:String \cap 代理审判员.xsd:String \cap 人民陪审员.xsd:String \cap 书记员.xsd:String \cap 检察员.xsd:String \cap Has 组织法院, 被告人 \sqsubseteq 被告人姓名.xsd:String \cap 被告人性别.xsd:String \cap 被告人年龄.xsd:String \cap 被告人出生年月.xsd:String \cap 被告人民族.xsd:String \cap 被告人住址.xsd:String \cap 被告人户籍地.xsd:String \cap 被告人文化水平.xsd:String \cap 被告人职业.xsd:String \cap 涉嫌罪名.xsd:String \cap 被羁押日期.xsd:String \cap 被逮捕日期.xsd:String \cap 现羁押地.xsd:String \cap 是否累犯.xsd:String \cap Has 法定代理人 \cap Has 辩护人, 起诉信息_被告人 \sqsubseteq 起诉信息_被告人姓名.xsd:String \cap 起诉信息_指控罪名.xsd:String, 辩护人 \sqsubseteq 辩护人姓名.xsd:String \cap 辩护人单位.xsd:String, 法定代理人 \sqsubseteq 法定代理人姓名.xsd:String \cap 法定代理人年龄.xsd:String \cap 法定代理人出生日期.xsd:String \cap 法定代理人职业.xsd:String \cap 法定代理人住址.xsd:String \cap 与被告人关系.xsd:String, 案情 \sqsubseteq 案情_被告人姓名.xsd:String \cap 犯案

地点.xsd:String \cap 被害人.xsd:String, 起因 \subseteq 事件.xsd:String \cap 先施暴方.xsd:String, 经过 \subseteq 施暴方式.xsd:String \cap 受害方.xsd:String \cap 受害人个数.xsd:String \cap 攻击次数 \cap Has 攻击部位 \cap Has 作案工具, 结果 \subseteq Has 伤情等级 \cap Has 从轻情节 \cap Has 从重情节, 物证 \subseteq 获取物证时间.xsd:String \cap 归属.xsd:String \cap 物品.xsd:String, 人证 \subseteq 证人证明时间.xsd:String \cap 证人涉案关系.xsd:String \cap 证人证明地点.xsd:String \cap 证人证明内容.xsd:String, 判决结果 \subseteq 罪名.xsd:String \cap 刑期.xsd:String \cap 其他附加.xsd:String \cap Has 依据法条, 法条 \subseteq 法条编号.xsd:String \cap 法条内容.xsd:String, 判决书 \subseteq 审理状态.xsd:String \cap 判决书编号.xsd:String \cap Has 起诉信息 \cap Has 审判机构 \cap Has 合议庭 \cap Has 被告人 \cap Has 案情 \cap Has 证据 \cap Has 判决结果 >

$TR = \{ \text{法院} \subseteq \text{法律机构}, \text{检察院} \subseteq \text{法律机构}, \text{被告人} \subseteq \text{法律角色}, \text{起诉信息} \subseteq \text{案情}, \text{被告人} \subseteq \text{被告人}, \text{辩护人} \subseteq \text{法律角色}, \text{法定代理人} \subseteq \text{法律角色}, \text{起因} \subseteq \text{案情}, \text{经过} \subseteq \text{案情}, \text{结果} \subseteq \text{案情}, \text{案情相关类} \subseteq \text{案情}, \text{攻击部位} \subseteq \text{案情相关类}, \text{作案工具} \subseteq \text{案情相关类}, \text{伤情等级} \subseteq \text{案情相关类}, \text{从轻情节} \subseteq \text{案情相关类}, \text{从重情节} \subseteq \text{案情相关类}, \text{物证} \subseteq \text{证据}, \text{人证} \subseteq \text{证据} \}$

刑事案件本体 XML 抽取模板

XML 标记模板如下所示，其中，由于 XML 文档的书写规则中涉及实体引用的约束，故将模式中的 “(?<=” 替换为 “(?<=”。

```
<?xml version="1.0" encoding="UTF-8"?>
<Judgement>
  <Entities_Judgement CName="刑事判决书">
    <Judge_number CName="判决书编号" pattern="\ ( \d{4} \)
[\u4E00-\u9FFF]+第\d+号\r">
    </Judge_number>
    <Judge_court CName="法院" pattern="[\u4E00-\u9FFF]+(?=
法院)法院\r">
    </Judge_court>
    <Judge_procuratorate CName="检察院" pattern="( ?&lt;=公
诉机关) [\u4E00-\u9FFF]+ ">
    </Judge_procuratorate>
    <Judge_fact_finder CName="检察员" pattern="( ?&lt;=(代
理)?检察员) [\u4E00-\u9FFF]+(?=(出庭|出席)) ">
```

```

</Judge_fact_finder>
<Judge_council CName="合议庭">
    <Judge_chief CName="审判长" pattern="( ?&lt;=审判长)
[\u4E00-\u9FFF]+\r">
        </Judge_chief>
        <Judge_member CName="审判员" pattern="( ?&lt;=审判员)
[\u4E00-\u9FFF]+\r">
            <Judge_member>
                <Judge_agent_chief CName="代理审判员" pattern="( ?&lt;=代
理审判长)[\u4E00-\u9FFF]+\r">
                    </Judge_agent_chief>
                    <Judge_agent_member CName="代理审判员" pattern=" ( ?&
lt;=代理审判员)[\u4E00-\u9FFF]+\r">
                        </Judge_agent_member>
                        <Judge_agent_member CName="人民陪审员" pattern=" ( ?&
lt;=人民陪审员)[\u4E00-\u9FFF]+\r">
                            </Judge_agent_member>
                            <Judge_clerk CName="书记员" pattern="( ?&lt;=书记员)
[\u4E00-\u9FFF]+\r">
                                </Judge_clerk>
                            </Judge_council>
                            <Judge_status CName="审理状态" pattern="现已审理终结">
                                </Judge_status>
                                <Roles CName="法律角色" pattern1="被告人.*\n" pattern2="(法
定)?代理人.*\n" pattern3="辩护人.*\n" type="Multiline">
                                    <Role_defendent CName="被告人" pattern="被告人.*">
                                        <Role_defendent_name CName="被告人姓名" pattern=" ( ?&
lt;=被告人)[\u4E00-\u9FFF]{2,4}(?= \ ( |, )">
                                            </Role_defendent_name>
                                            <Role_defendent_sex CName="被告人性别" pattern=" ( ?&
lt;=, ) [男女](?=, )">
                                                </Role_defendent_sex>
                                                <Role_defendent_age CName="被告人年龄" pattern=" ( ?&

```

```

lt;=, )\d{1,2}岁(?, )">
    </Role_defendent_age>
    <Role_defendent_birthdate CName="被告人出生年月" pattern=
"(?&lt;=, )\d{4}年\d{1,2}月\d{1,2}日(?=出生于)">
    </Role_defendent_birthdate>
    <Role_defendent_birthplace CName="被告人出生地" pattern=
"(?&lt;=出生于)[\u4E00-\u9FFF]+(?, )">
    </Role_defendent_birthplace>
    <Role_defendent_nation CName="被告人民族" pattern=
"(?&lt;=, )[\u4E00-\u9FFF]+族(?, )">
    </Role_defendent_nation>
    <Role_defendent_education CName="被告人文化水平" pattern=
"(?&lt;=, )[\u4E00-\u9FFF]+文化(?, )">
    </Role_defendent_education>
    <Role_defendent_career CName="被告人职业" pattern=
"(?&lt;=文化, )[\u4E00-\u9FFF]+(?, )">
    </Role_defendent_career>
    <Role_defendent_address CName="被告人住址" pattern=
"(?&lt;=, 住|暂住).*?(?, |。)">
    </Role_defendent_address>
    <Role_defendent_residence CName="被告人户籍地" pattern=
"(?&lt;=, 户籍所在地|户籍地为).*?(?, |。|)">
    </Role_defendent_residence>
    <Role_defendent_crime CName="涉嫌罪名" pattern= "(?&
lt;=因涉嫌犯) [\u4E00-\u9FFF]+ (?=于\d{4}年\d{1,2}月\d{1,2}日
被)">
    </Role_defendent_crime>
    <Role_defendent_detained CName="被羁押日期" pattern=
"(\d{4}|同)年\d{1,2}月\d{1,2}日(?=被羁押)">
    </Role_defendent_detained>
    <Role_defendent_arrested CName="被逮捕日期" pattern=
"(?&lt;=于) \d{4}年\d{1,2}月\d{1,2}日(?=被逮捕)">
    </Role_defendent_arrested>

```

```

        <Role_defendent_detainplace CName="现羁押地" pattern=
"(?&lt;=现羁押(于|在)) [\u4E00-\u9FFF]+(?=。|,)">
        </Role_defendent_detainplace>
        <Judge_punish_lighter CName="从轻情节" pattern="自首
|积极赔偿|投案|积极救治|从犯|如实供述|揭发|认罪悔罪|未成年">
        </Judge_punish_lighter>
        <Judge_punish_heavier CName="从重情节" pattern="逃逸
|累犯.....">
        </Judge_punish_heavier>
    </Role_defendent>
    <Role_defender CName="辩护人" pattern="辩护人.*">
        <Role_defender_name CName="辩护人姓名" pattern="( ?&
lt;=(指定|法定代理人暨)?辩护人)([\u4E00-\u9FFF]{2,3}、?)+(?=, )">
        </Role_defender_name>
        <Role_defender_unit CName="辩护人单位" pattern="( ?&lt;=
(指定|法定代理人暨)?辩护人([\u4E00-\u9FFF]{2,3}、?), )[\u4E00-\
u9FFF]+(?=。)">
        </Role_defender_unit>
    </Role_defender>
    <Role_agent CName="被告人" pattern="(法定)?代理人.*">
        <Role_agent_name CName="代理人姓名" pattern="( ?&lt;=
法定代理人暨辩护人)([\u4E00-\u9FFF]{2,3}、?)+(?=, | )">
        </Role_agent_name>
        <Role_agent_relation CName="与被告关系" pattern="( ?&lt;=
\ ( )[\u4E00-\u9FFF]+(?=\ ) )">
        </Role_agent_relation>
        <Role_agent_age CName="代理人年龄" pattern="( ?&lt;=, )\
d{1,2}岁(?=, )">
        </Role_agent_age>
        <Role_agent_birthdate CName="代理人出生年月" pattern=
"\d{4}年\d{1,2}月\d{1,2}日">
        </Role_agent_birthdate>
        <Role_agent_nation CName="代理人民族" pattern="( ?&lt;=, )

```



```

[\u4E00-\u9FFF]+族(?, )">
    </Role_agent_nation>
    <Role_agent_career CName="代理人职业" pattern="( ?&lt;=
(文化, )|(族, ))[\u4E00-\u9FFF]+(?, )">
    </Role_agent_career>
    <Role_agent_address CName="代理人住址" pattern="
( ?&lt;=, 住|暂住).*?(?=, |。| )">
    </Role_agent_address>
</Role_agent>
<Charge CName="起诉书" pattern1=[\u4E00-\u9FFF]+院以[\u4
E00-\u9FFF]+[\d{4}]\d{3,6}号.*\r pattern2=.*院(起诉书)?指
控,.*\r pattern3=针对(上述)?指控(的)?(上述)?(事实)?, 公诉(机关|
人)(当庭|向法庭).*\r pattern4=.*(((被告人[\u4E00-\u9FFF]{2,4}的|
其)辩护人(的辩护意见是|认为))|(被告人[\u4E00-\u9FFF]{2,4}及辩护人均
认为.*))" type="Multiline">
    <Charge_info CName="起诉书基本信息" pattern="[\u4E00-
\u9FFF]+院以[\u4E00-\u9FFF]+ \d{4}\d{3,6}号.*\r">
    <Charge_procuratorate CName="检察院" pattern="^[ \u4E00-
\u9FFF]+院(=?以)">
    </Charge_procuratorate>
    <Charge_No CName="起诉书编号" pattern="( ?&lt;=以 )
[\u4E00-\u9FFF]+ \d{4} \d{3,6}、?)+号">
    </Charge_No>
    <Charge_date CName="提起公诉日期" pattern="( ?&lt;=
于)\d{4}年\d{2}月\d{2}日(=?向本院提起公诉)">
    </Charge_date>
    <Charge_desendent CName="起诉被告人">
    <Charge_desendent_name CName="起诉信息_被告人姓名
" pattern="( ?&lt;=被告人)([\u4E00-\u9FFF]{2,5}、?)+(=?犯
[\u4E00-\u9FFF]+罪, )">
    </Charge_desendent_name>
    <Charge_desendent_crime CName="起诉信息_指控罪名"
pattern="( ?&lt;=犯)([\u4E00-\u9FFF]{2,5}、?)+罪(=? , )">

```

```

        </Charge_desendent_crime>
    </Charge_desendent>
    <Charge_content CName="指控内容" pattern=".*院(起诉书)?
指控,.*\r">
        <Charge_case_time CName="指控_案发时间" pattern=
"\d{4}年\d{1,2}月\d{1,2}日(凌晨,|\d{1,2}时许,)">
            </Charge_case_time>
            <Charge_case_place CName="指控_案发地点" pattern="(在
|至).*(室|附近|内|村|上)">
                </Charge_case_place>
                <Charge_case_victim CName="指控_被害人姓名" note="结
合句法分析">
                    </Charge_case_victim>
                    <Charge_case_victim_number CName="指控_被害人个数"
note="Count(被害人姓名)">
                        </Charge_case_victim_number>
                        <Charge_case_cause CName="指控_案发起因" pattern="因
[\u4E00-\u9FFF]+(?:=。|,)" note="结合句法分析">
                            </Charge_case_cause>
                            <Charge_case_process CName="指控_案件经过" pattern=
".*院(起诉书)?指控,.*\r">
                                </Charge_case_process>
                                <Charge_case_result CName="指控_案件结果" pattern=
"(造成|致).*(伤|亡)">
                                    </Charge_case_result>
                                </Charge_content>
                                <Charge_conclusion CName="指控总结" pattern="针对(上述)?
指控(的)?(上述)?(事实)?, 公诉(机关|人)(当庭|向法庭).*\r">
                                    <Charge_evidence CName="指控_证据" pattern="针对(上
述)?指控(的)?(上述)?(事实)?, 公诉(机关|人)(当庭|向法庭).*等证据"
note="结合句法分析">
                                        </Charge_evidence>
                                        <Charge_law CName="指控_法条依据" pattern="(?:&lt;=其

```

```

行为触犯了).*规定">
    </Charge_law>
    <Charge_crime CName="指控_罪责" pattern="( ?&lt;!=已构成|应当以|应以)([\\u4E00-\\u9FFF]+罪、 ?)+ ">
    </Charge_crime>
    <Charge_reference CName="指控_参考" pattern="( ?&lt;!=应当)((从轻(或减轻)?)|从重)处罚">
    </Charge_reference>
    </Charge_conclusion>
    <Charge_defend CName="辩护意见">
        <Charge_defend_bydefendent CName="被告意见" pattern=
"(在法庭审理过程中, )?( (.*(?=(被告人[\\u4E00-\\u9FFF]{2,4}的|其)辩护人(的辩护意见是|认为)))|(被告人[\\u4E00-\\u9FFF]{2,4}及辩护人均认为.*)) ">
        </Charge_defend_bydefendent>
        <Charge_defend_bydefender CName="辩护人意见" pattern=
"((被告人[\\u4E00-\\u9FFF]{2,4}的|其)辩护人(的辩护意见是|认为).*)|(被告人[\\u4E00-\\u9FFF]{2,4}及辩护人均认为.*)) ">
        </Charge_defend_bydefender>
    </Charge_defend>
</Charge>
<Case CName="案情" pattern="经(本院)?审理查明, .*\\r 被告人.*(归案|投案)。" type="Multiline">
    <Case_time CName="案发时间" pattern="\\d{4}年\\d{1,2}月\\d{1,2}日(晚|\\d{1,2}时许)?">
    </Case_time>
    <Case_place CName="案发地点" pattern="(在|位于).*(附近|外|内|侧)">
    </Case_place>
    <Case_desendent_name CName="案情_被告人姓名" pattern=
"( ?&lt;!=被告人)[\\u4E00-\\u9FFF]{2,4}" note="结合句法分析">
    </Case_desendent_name>
    <Case_victim_name CName="案情_被害人姓名" pattern=" ( ?&

```

```

lt;=被害人)[\u4E00-\u9FFF]{2,4}" note="结合句法分析">
    </Case_victim_name>
    <Case_victim_number CName="案情_被害人个数" note="count
(被害人姓名)">
        </Case_victim_number>
        <Case_affair CName="事件" pattern="因[\u4E00-\u9FFF]+
(?.。|,|), " note="结合句法分析">
            </Case_affair>
            <Case_violenceway CName="施暴方式" pattern="厮打|互殴|殴
打|猛刺|击打" note="结合句法分析">
                </Case_violenceway>
                <Case_attackpart CName="攻击部位" pattern="((胸(腹)?部)|
腹部|(头(面)?部)|四肢部|背部|肋部|颈项部|(左|右)?肩部|身体)(等)?"
note="结合句法分析">
                    </Case_attackpart>
                    <Case_attacktimes CName="攻击次数" note="结合句法分析">
                        </Case_attacktimes>
                        <Case_attacktool CName="作案工具" pattern="((?&lt;=持)
([\u4E00-\u9FFF]+、?)+(?.猛刺|朝|等))|(?&lt;=被)[\u4E00-
\u9FFF]+(?.(多次)?击打)">
                            </Case_attacktool>
                            <Case_result CName="伤情等级" pattern="致.*(死亡|(轻(微)?
伤)|重伤)">
                                </Case_result>
                            </Case>
                            <Proof CName="证据" pattern="(?.&lt;=本院予以确认的下列证据证实
|上述事实, 有下列证据证实: )(.*\r)+(?.本院认为)" type="Multiline">
                                <Proof_person CName="证人证言">
                                    <Proof_per CName="证人_证言" pattern="(\d{1,2}、)?(证人|共同
犯罪人|被害人)([\u4E00-\u9FFF]+、?)+((证实: )|证明|证言).*">
                                        <Proof_pers_name CName="证人姓名" pattern="(?.&lt;=
(证人|共同犯罪人|被害人))([\u4E00-\u9FFF]+、?)(?.证言(分别)?证实|
证明|证言)">

```

```

        </Proof_per_name>
        <Proof_per_time CName=" 证人证明时间 " pattern=
"\d{4}年\d{1,2}月\d{1,2}日(晚|\d{1,2}时许|晨)?">
        </Proof_per_time>
        <Proof_per_place CName="证人证明地点" note="结合句
法分析">
        </Proof_per_place>
        <Proof_per_relation CName="与被告关系" pattern=
"(?&lt;=( 证人 | 共同犯罪人 | 被害人 )[\u4E00-\u9FFF]+\ ( )([\u4E00-
\u9FFF]+、 ?)+(?= \ ))">
        </Proof_per_relation>
        <Proof_per_content CName="证人证明内容" pattern=
"(?&lt;=( 证明 | 证实 | 证言 ): ) . * ">
        </Proof_per_content>
    </Proof_per>
<Proof_person>
<Proof_thing CName="物证">
    <Proof_object CName=" 证 物 " pattern="[\u4E00-
\u9FFF]+( 鉴定书 | 记录 | 登记表 | 照片 | 报告 | 笔录 | 餐饮单 | ( 户籍 ( 材料 ) ? 证明 ) |
过程 | 勘查笔录 ) 证实 . * ">
    <Proof_object_type CName="物证类别" pattern="[\u4E00-
\u9FFF]+( ? = 证实 ) ">
    </Proof_ object _type>
    <Proof_ object _content CName="证明内容" pattern=
"( ? &lt; = 证实 ) . * ">
    </Proof_ object _content>
</Proof_ object >
</Proof_thing>
</Proof>
<Judge CName="判决" pattern="本院认为， ( . * \r ) + ( ? = 审判长 ) "
type="Multiline">
    <Judge_reason CName="判决依据" pattern="本院认为， . * ( ? = 判
决如下 ) ">

```

```

        <Judge_reason_case CName="案情依据" pattern="本院认为, .*(?=据此)">
            </Judge_reason_case>
            <Judge_reason_law CName="法条依据" pattern="(?!<=据此, ).*(?=判决如下)" note="结合句法分析">
                </Judge_reason_law>
            </Judge_reason>
            <Judge_result CName="判决结果" pattern="被告人.*">判决结果_判决书编号
                <Judge_result_defendence CName="被告人" pattern="(?!<=被告人)[\u4E00-\u9FFF]{2,4}(?=犯)">
                    </Judge_result_defendence>
                    <Judge_result_crime CName="罪名" pattern="(?!<=犯)([\u4E00-\u9FFF]+罪、?)+">
                        </Judge_result_crime>
                        <Judge_result_crimeperiod CName="刑期" pattern="判处.*(?!<=剥夺))?">
                            </Judge_result_crimeperiod>
                            <Judge_result_others CName="其他附加" pattern="剥夺政治权利([\u4E00-\u9FFF]+年)|终身">
                                </Judge_result_others>
                            <Judge_result>
                        <Judge>
                            </Entities_Judgement>
                            <Relations_Judgement>
                                <Relation_JudgeProcue RName="Has 起诉书">
                                    <Source CName="刑事判决书"></Source>
                                    <Domain CName="起诉书"></Domain>
                                </Relation_JudgeProcue>
                                <Relation_JudgeDefendent RName="Has 被告人">
                                    <Source CName="刑事判决书"></Source>
                                    <Domain CName="被告人"></Domain>
                                </Relation_JudgeDefendent>

```

```

<Relation_JudgeCase RName="Has 案情">
  <Source CName="刑事判决书"></Source>
  <Domain CName="案情"></Domain>
</Relation_JudgeCase>
<Relation_JudgeProof RName="Has 证据">
  <Source CName="刑事判决书"></Source>
  <Domain CName="证据"></Domain>
</Relation_JudgeProof>
<Relation_JudgeJudge RName="Has 判决">
  <Source CName="刑事判决书"></Source>
  <Domain CName="判决"></Domain>
</Relation_JudgeJudge>
<Relation_defend RName="Has 辩护人">
  <Source CName="被告人"></Source>
  <Domain CName="辩护人"></Domain>
</Relation_Defend>
<Relation_Agent RName="Has 代理人">
  <Source CName="被告人"></Source>
  <Domain CName="代理人"></Domain>
</Relation_Agent>
<Relation_CourtCouncil RName="Has 合议庭">
  <Source CName="法院"></Source>
  <Domain CName="合议庭"></Domain>
</Relation_CourtCouncil>
</Relations_Judgement>
</Judgement>

```

实验数据

训练集

(随机化后 1~100 号案例的节选)

序号	案例号	事因	其他原因	酒后	先施暴方	其他行为	打	刺	趁人不备	其他施暴方式	受害人数	拳脚	钝器	锐器	其他	胸部	腹部	四肢	身体	头部	面部	眼部	颈部	伤情等级	多人作案	被害人负有一定责任	未成年	限制刑事责任能力人	自首	积极救治	认罪	悔罪	积极赔偿	得到谅解	达成民事赔偿协议	被害人方面要求从轻	其他从轻条件	累犯	主犯	劳教人员	刑期	缓刑	随机排序
44	111_0166126_1	1	A8	0	B2	C4	1	0	0	1	1	0	0	0	0	0	0	0	0	1	0	0	D4	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0	E1	1	0.615164
125	111_0197943_2	1	A8	0	B2	C4	1	0	0	1	2	0	1	0	1	0	0	0	0	1	1	0	D9	0	0	0	0	0	0	0	1	1	1	0	0	1	0	0	0	0	E4	3	0.031386
78	111_0179804_1	1	A8	0	B3	C4	1	1	0	0	2	1	0	1	0	0	0	0	0	0	1	0	D4	0	0	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	E1	1	0.141137
55	111_0172346_1	1	A8	0	B2	C4	1	0	0	1	1	0	0	0	0	0	0	0	0	0	1	0	D4	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0	E1	1	0.298838	
52	111_0170019_1	1	A8	0	B2	C4	1	0	0	1	0	1	0	0	0	0	0	0	0	1	0	0	D4	0	0	0	0	0	1	0	0	0	1	0	0	0	0	0	0	E1	1	0.566409	
130	111_0200454_2	1	A8	0	B2	C4	1	0	0	1	1	0	0	0	0	0	0	0	0	0	1	0	D9	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	E4	4	0.92868
99	111_0185265_2	1	A8	0	B2	C4	1	0	0	1	0	1	0	0	0	0	0	0	0	0	1	0	D4	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0	E1	1	0.361321	
61	111_0174890_1	1	A8	0	B2	C4	1	0	0	1	1	0	0	0	1	0	0	0	0	0	0	0	D4	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	E1	0.7	0.136544	
131	111_0200682_2	1	A8	0	B3	C4	1	0	0	1	1	0	0	0	0	0	0	0	0	1	0	0	D4	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	E1	1	0.668193	

续表

序号	案例号	事因	其他原因	酒后	先施暴方	其他行为	打	刺	趁人不备	其他施暴方式	受袭人数	拳脚	钝器	锐器	其他	胸部	腹部	四肢	身体	头部	面部	眼部	颈部	伤情等级	多人作案	被害人负有一定责任	未成年	限制刑事责任能力人	自首	积极救治	认罪	悔罪	积极赔偿	得到谅解	达成民事赔偿协议	被害人方面要求从轻	其他从轻条件	累犯	主犯	劳教人员	刑期	缓刑	随机排序	
26	111_0158244_1	1	A8	0	B2	C4	0	1	0	0	1	0	0	1	0	0	0	0	1	0	0	0	0	D7	0	0	0	0	0	0	1	0	1	0	0	0	0	0	0	0	E4	4	0	0.611964
74	111_0176917_1	1	A3	0	B2	C4	0	1	1	0	1	0	0	1	0	0	1	0	0	0	0	0	D9	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0	E4	5	0	0.630974
76	111_0178390_1	1	A8	0	B2	C4	1	0	0	0	1	1	0	0	0	0	0	0	0	0	1	0	D4	0	0	0	0	0	0	1	1	0	1	0	0	0	0	0	0	E1	0.8	1	0.928746	
43	111_0165443_1	1	A8	0	B3	C4	1	0	0	0	1	1	0	0	0	0	0	1	0	0	1	0	D4	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	E1	1	0	0.669602	
77	111_0178811_1	1	A8	0	B2	C4	1	0	0	0	1	1	0	0	0	0	0	0	0	0	1	0	D4	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	E1	0.7	1	0.603539	
14	110_0171624_1	1	A8	0	B2	C4	0	1	0	0	1	0	0	1	0	1	0	0	0	1	0	0	D11	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	E5	12	0	0.344702	
20	111_0154951_1	1	A8	0	B2	C4	0	1	0	0	1	0	0	1	0	1	1	0	0	0	0	0	D6	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	E1	1	1	0.175881	
6	120_0238899_2	0	A4	0	B2	C1	1	0	0	0	2	0	1	1	0	0	0	0	1	0	0	0	D11	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	E4	9	0	0.545491

测试集

(随机化后 101~150 号案例的节选)

序号	案例号	事因	其他原因	酒后	先施暴方	其他行为	打	刺	趁人不备	其他施暴方式	受害人数	拳脚	钝器	锐器	其他	胸部	腹部	四肢	身体	头部	面部	眼部	颈部	伤情等级	多人作案	被害人负有一定责任	未成年	限制刑事责任能力人	自首	积极救治	认罪	悔罪	积极赔偿	得到谅解	达成民事赔偿协议	被害人方面要求从轻	其他从轻条件	累犯	主犯	劳教人员	刑期	缓刑	随机排序	
36	111_0163485_1	1	A8	0	B2	C4	1	0	0	0	1	0	1	0	0	0	0	0	1	0	0	0	0	D4	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0	E1	1.5	0	0.874214
27	111_0158246_1	1	A8	0	B2	C4	1	0	0	0	1	1	0	0	0	0	0	0	0	0	1	0	D4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	E1	1	0	0.772893
129	111_0200061_3	0	A5	0	B2	C4	1	0	0	0	1	0	0	1	0	0	0	0	0	1	0	0	D11	0	0	0	0	1	0	0	0	1	1	0	0	0	0	0	0	0	E4	4	0	0.597574
33	111_0161920_1	1	A8	0	B3	C4	1	0	0	0	1	1	0	0	0	0	0	0	1	0	0	0	D4	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0	E1	0.7	1	0.004961	

续表

序号	案 例 号	事 因	其 他 原 因	酒 后	其 他 行 为	打	趁 人 不 备	其 他 施 暴 方 式	受 害 人 数	拳 脚	钝 器	锐 器	其 他	胸 部	腹 部	四 肢	身 体	头 部	面 部	眼 部	颈 部	伤 情 等 级	多 人 作 案	被 害 人 负 有 一 定 责 任	未 成 年	限 制 刑 事 责 任 能 力 人	自 首	积 极 救 治	认 罪	悔 罪	积 极 赔 偿	得 到 谅 解	达 成 民 事 赔 偿 协 议	被 害 人 方 面 要 求 从 轻	其 他 从 轻 条 件	累 犯	主 犯	劳 教 人 员	刑 期	缓 刑	随 机 排 序		
57	111_0173683_1	1	A8	0	B2	C4	1	0	0	1	0	0	0	0	0	1	0	0	0	0	0	D4	0	0	0	0	0	1	0	1	0	1	0	0	0	0	0	0	E1	1	0	0.680172	
56	111_0172723_1	1	A8	0	B2	C4	0	1	0	3	0	0	1	0	1	1	0	1	0	0	0	D9	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	E4	4	0	0.770327	
111	111_0190974_2	1	A8	0	B3	C4	1	0	0	1	0	0	0	0	0	0	0	1	0	0	D4	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	E1	1	1	0.16023	
119	111_0195546_2	1	A2	0	B2	C1	1	0	0	2	0	1	0	0	0	0	0	1	0	0	D5	0	0	0	1	0	1	0	1	0	1	0	1	0	0	0	0	0	E1	0.5	1	0.789718	
95	111_0183140_1	1	A8	0	B2	C4	0	1	0	1	0	0	1	1	0	1	0	1	0	0	D9	0	1	0	0	1	0	0	0	1	0	0	1	0	0	0	0	0	E4	6	0	0.938743	
83	111_0180272_1	1	A8	0	B2	C4	1	0	0	1	0	0	0	0	0	0	0	1	0	0	D4	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	E1	1	1	0.900377	
80	111_0179844_1	1	A8	0	B2	C1	1	0	0	2	0	1	0	0	0	0	0	1	0	0	D4	0	0	0	1	0	0	0	1	1	0	0	1	0	0	0	0	0	E1	0.5	1	0.753661	
114	111_0191838_1	1	A8	0	B2	C4	1	0	0	1	0	1	0	0	0	1	0	0	0	0	D4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	E1	0.7	0	0.994715	
51	111_0169529_1	1	A8	0	B2	C4	0	1	0	1	0	0	1	0	0	1	0	0	0	0	D9	0	0	0	0	0	0	0	0	1	0	1	0	1	0	0	0	0	E4	3.5	0	0.465069	
75	111_0177957_1	1	A8	0	B2	C2	1	0	0	2	0	1	1	0	1	0	1	0	0	0	D4	0	0	0	0	0	0	0	1	0	1	0	1	0	0	0	0	0	E1	1	0	0.147202	
68	111_0176455_2	1	A8	0	B2	C4	1	0	0	1	0	0	0	0	0	0	0	1	0	0	D4	0	0	0	0	0	0	0	0	1	1	0	1	0	0	0	0	0	E1	0.8	1	0.399839	
45	111_0166697_1	1	A8	0	B2	C4	1	0	0	1	0	0	0	0	0	0	0	1	1	0	D4	0	0	0	0	0	1	0	1	1	1	0	1	0	0	0	0	0	E1	0.7	1	0.85123	
128	111_0199967_3	1	A8	0	B2	C4	0	1	0	2	0	0	1	0	1	0	1	0	0	0	D9	0	0	0	0	0	1	0	0	0	0	1	1	0	0	0	0	0	0	E1	1	0	0.682557
110	111_0190753_1	1	A8	0	B2	C4	1	0	0	1	0	0	0	0	0	0	0	1	0	0	D4	0	0	0	0	0	0	0	0	1	0	1	0	1	0	0	0	0	E1	1	1	0.389795	
88	110_0205209_1	1	A8	1	B3	C4	0	1	0	0	1	0	1	0	0	0	0	0	0	0	D11	0	0	0	0	0	0	0	1	0	1	0	0	0	1	1	0	0	E5	15	0	0.12651	
38	111_0164203_1	1	A8	0	B2	C4	1	0	0	1	0	0	1	0	0	0	0	1	0	0	D4	0	0	0	0	0	0	0	1	0	0	0	1	1	0	0	0	0	E1	0.5	0	0.379086	
7	110_0148767_1	1	A6	0	B2	C4	1	0	0	1	0	1	0	0	0	0	1	0	0	0	D11	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	E5	12	0	0.885736	
147	121_0128707_2	1	A8	0	B2	C4	1	0	1	1	0	0	0	0	0	0	1	0	1	0	D6	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	E1	2	0	0.860131	
94	111_0182618_1	1	A8	0	B3	C4	1	0	0	1	0	0	0	0	0	0	0	0	1	0	D4	0	0	0	0	0	0	0	1	1	0	0	1	0	0	0	0	0	E1	1	1	0.69665	
144	121_0118114_2	1	A8	0	B3	C4	1	0	0	1	0	0	0	0	0	0	0	0	1	0	D4	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	E1	0.8	0	0.140124	

后 记

本书是作者攻读博士学位期间的研究成果。本书在研究期间，依托了国家自然科学基金项目《基于 CAS 的焦点企业核型结构产业集群创新网络演化机理研究》(70972115) (2009—2012)、北京市哲学社会科学规划项目《通过技术并购促进北京市传统优势企业转型升级的路径研究》(12JGB070) 和横向项目《司法审判领域数据挖掘研究》(2012—2014)，本书所有学术成果及学术观点属于这些课题成果之一。值此图书出版之际，感谢在本书撰写过程中帮助和关心我的各位导师、朋友。

感谢我的导师张永安教授。和张教授求学的四年多时间里，张教授用博学的精神、严谨的态度和持久的方法，让我深深感受到导师教书的高尚情操，每两周一次的问题研讨、硕博论坛和每到一个阶段对我的及时指导与鼓励，都给我留下了深刻印象，使我在整个学习过程中不断感受到激励，不断增强战胜自我和克服困难的勇气。张教授教学更注重育人，当发现我的问题时，会用长者对晚辈的方式给予指出，谆谆教诲我要学会“控制”，发挥自己的长处，弥补自己的不足，教我学会处理问题的方式、方法。张教授用行动时时感染着我，记得每次论文修改，我下班后发给张老师，总会在凌晨就收到张老师的指导，自己常常会想，张教授为了我和我的同门，得付出多少个不眠之夜。因为我是在

职学习，没有在校生那么多精力，张老师为了我能如期完成学业，总会在每一周给我指出下周应该完成的任务，用他的方法给我以鼓励和帮助，给我信心。在此，向我尊敬的导师致以最真挚的谢意！

感谢刘超教授、黄鲁成教授、肖创柏教授、翟东升教授、杨松令教授、蒋国瑞教授、韩福荣教授、顾力刚教授、徐艳梅教授、官建成教授、韩柏棠教授、王刊良教授、陈向东教授和郑海航教授，感谢各位老师对本书的撰写提出了宝贵修改意见，让我能够顺利完成全书的创作。

感谢我的师弟、师妹和研究团队对本书完成付出的辛勤劳动，谢谢你们给予我的支持和无私帮助。

感谢北京市高级人民法院领导对本研究给予的时间和数据资源上的大力支持。

参考文献

- [1] 王磊, 胡鞍钢. 结构、能力与机制: 中国决策模式变化的实证分析. 探索与争鸣, 2010, 6: 3-7.
- [2] 刘海学. 基于语义标注的元数据自动建立及其相关技术研究[D]. 上海: 华东师范大学, 2010.
- [3] Cardie C. Empirical Methods in Information Extraction[J]. AI Magazine, 1997, 18(4):65-78.
- [4] 荆涛等. 中文网页语义标注: 由句子到 RDF 表示[J]. 计算机研究与发展, 2008, 45 (7): 1221-1231.
- [5] Dill S, Tomlin J, et al. SemTag and Seeker: Bootstrapping the Semantic Web Via Automated Semantic Annotation[C]. Proc of the 12th Int'l Conf on World Wide Web. New York: ACM, 2003:178-186.
- [6] Alani H, Kim S, Millard D, et al. Automatic Ontology-Based Knowledge Extraction from Web Documents[J]. Intelligent Systems, 2003, 18(1):13-21.
- [7] Schutz A, Buitelaar P. RelExt: A Tool for Relation Extraction from Text in Ontology Extension[C]. Proc of the 4th Int'l Semantic Web Conf (ISWC). Berlin: Springer, 2005:593-606.
- [8] Tenier S, Toussaint Y, Napoli A, et al. Instantiation of Relations for Semantic Annotation[C]. Proc of Web Intelligence 2006. Los Alamitos: IEEE Computer Society, 2006:463-472.

- [9] Gruber. A Translation Approach to Portable Ontology Specifications[J]. Knowledge Acquisition, 1993, 5(2):199-220.
- [10] Laurens Mommers, Aernout Schmidt, Eduard Oskamp. Controversies in the Ontology and Law Debate[C]. In: Proceedings of the 1st International Workshop on Legal Ontologies, eds. P. Visser, and R. Winkels, 1-5, Melbourne, Australia, 1997.
- [11] Adam Wyner. An Ontology in OWL for Legal Case-Based Reasoning[J]. Artificial Intelligence and Law, 2008, 16(4): 361-387.
- [12] Valente A, Breuker J. Ontologies: The missing link between legal theory and AI & Law[C]. JURIX, 1994, 94: 139-149.
- [13] Antoniou G, Van Harmelen F. Semantic Web Primer[M]. London, England: the MIT Press, 2004.
- [14] McGuinness D L. Ontologies Come of Age[J]. Spinning the Semantic Web: Bringing the World Wide Web to its full potential, 2005: 171.
- [15] Casanovas P, Noriega P, Bourcier D, Galindo F. Trends in Legal Knowledge—The Semantic Web and the Regulation of Electronic Social Systems[M]. European Press Academic Publication, Florence, Italy, 2007.
- [16] Breuker J, Valente A, Winkels R. Legal Ontologies in Knowledge Engineering and Information Management[J]. Artificial Intelligence and Law, 2004, 12(4): 241-277.
- [17] McCarty L T. A Language for Legal Discourse I. Basic Features[C]. In: Proceedings of the 2nd International Conference on Artificial Intelligence and Law. ACM, 1989: 180-189.
- [18] Fellbaum. WordNet: An Electronic Lexical Database[M]. Cambridge, MA: MIT Press, 1998.
- [19] Gangemi A, Sagri M T, Tiscornia D. Metadata for Content Description in Legal Information[C]. In: Proceedings of the ICAIL 2003 Workshop on Legal

Ontologies & Web Based Legal Information Management, 2003.

[20] Dini L, Peters W, Liebwald D, et al. Cross-lingual Legal Information Retrieval Using a WordNet Architecture[C]. In: Proceedings of the 10th International Conference on Artificial Intelligence and Law. ACM, 2005: 163-167.

[21] Van Kralingen. Frame-based Conceptual Models of Statute Law[D]. Leiden University, 1995.

[22] Verheij B, Hage J. States of Affairs, Events, and rules: an Abstract Model of the Law[J]. Legal Knowledge Based Systems, 1997: 3-19.

[23] Mommers L. Applied Legal Epistemology: Building a Knowledge-based Ontology of Law. Leiden University, 2002.

[24] Wyner A, Mochales-Palau R, Moens M F, et al. Approaches to Text Mining Arguments from Legal Cases[M]. Semantic Processing of Legal Texts. Springer Berlin Heidelberg, 2010: 60-79.

[25] Chou S, Hsing T P. Text Mining Technique for Chinese Written Judgment of Criminal Case[M]. Intelligence and Security Informatics. Springer Berlin Heidelberg, 2010: 113-125.

[26] McCarty L T. A Language for Legal Discourse I. Basic Features[C]. In: Proceedings of the 2nd International Conference on Artificial Intelligence and Law. ACM, 1989: 180-189.

[27] Valente A. Legal Knowledge Engineering: A Modelling Approach[D]. Netherlands, University of Amsterdam, 1995.

[28] Kralingen V. Frame-based Conceptual Models of Statute Law[C]. Computer/Law Series, No. 16, Kluwer Law International, The Netherlands: The Hague, 1995.

[29] Varzi A C, Vieu L. Core Concepts of Law: Taking Common Sense

Seriously[C]. Formal Ontology in Information Systems: Proceedings of the Third International Conference (FOIS-2004), [November 3-6, Turin, Italy]. IOS Press, 2004: 210.

[30] Hoekstra R. Ontology Representation Design Patterns and Ontologies that Make Sense[C]. In: Proceedings of the 2009 Conference on Ontology Representation: Design Patterns and Ontologies that Make Sense. Ios Press, 2009: 1-236.

[31] Peters W, Sagri M T, Tiscornia D. The Structuring of Legal Knowledge in LOIS[J]. Artificial Intelligence and Law, 2007, 15(2): 117-135.

[32] Asaro C, Biasiotti M A, Guidotti P, Papini M, Sagri M T, Tiscornia D[EB/OL]. <http://www.lri.jur.uva.nl/~win-kels/LegOnt2003/Asaro.pdf>, 2013-05-22.

[33] Lame G. Constructing an IR-oriented Legal Ontology[EB/OL]. <http://www.lri.jur.uva.nl/jurix2001/papers/lame.pdf>, 2010-05-22.

[34] Giereth M, Ertl T. Visualization Enhanced Semantic Wikis for Patent Information[C] //Information Visualisation, 2008. IV'08.12th International Conference. IEEE, 2008: 185-190.

[35] Sylvie Despres, Sylvie Szulman. Merging of Legal Micro-Ontologies from European Directives[J]. Artificial Intelligence and Law, 2007(15):187-200.

[36] Cestnik B, Kern A, Modrijan H. Semi-automatic Ontology Construction for Improving Comprehension of Legal Documents[M]. Springer Berlin Heidelberg, 2008.

[37] Lame G. Using NLP Techniques to Identify Legal Ontology Components: Concepts and Relations[M]//Law and the Semantic Web. Springer Berlin Heidelberg, 2005: 169-184.

[38] Kingston J, Schafer B, Vandenberghe W. Towards a Financial Fraud Ontology: A Legal Modelling Approach[J]. Artificial Intelligence and Law, 2004, 12(4): 419-44.

- [39] 李子叶, 等. 基于模糊集贴近度的文本信息分类器分辨率改进算法研究. 管理工程学报, 2009, 23 (1).
- [40] 王丽珍. 一种基于语义贴近度的抽象归纳法. 计算机学报, 2000, 23 (10).
- [41] 李蕊, 等. 基于语义相关和概念相关的自动分类方法研究. 计算机工程与应用, 2003, 12: 106-109.
- [42] 程少川, 等. 基于本体系统的知识管理人工智能研究. 管理科学学报, 2003, 6 (2): 17-22.
- [43] 王娜, 李云松. 基于概念格的文本挖掘. 计算机技术与发展, 2006, 16 (1): 113-116.
- [44] 汪庆, 等. 连续特征离散化方法综述. 第五届管理科学与工程论坛论文集《2008 中国发展进程中的管理科学与工程》, 2008 (2): 1091-1096.
- [45] 马志锋, 等. 决策表中规则获取的不确定性研究. 控制与决策, 2000, 15 (6): 703-707.
- [46] 陆光辉, 肖人彬. 基于粗集理论和神经网络的集成化数据挖掘方法研究. 小型微型计算机系统, 2002, 23 (5): 552-557.
- [47] 徐亚娟. 基于公安业务信息的文本挖掘技术研究与实现. 浙江大学硕士论文, 2008.
- [48] 孙金年. 知识的存在形式. 南京大学学报, 2003, 40 (1): 90-97.
- [49] Sharjeel Imtiaz, Azmat Hussain. Using Agents for Unification of Information Extraction and Data Mining. IEEE, May 2005, 197-200.
- [50] Dozier C, Jackson P. Mining Text for Expert Witnesses [J]. IEEE Software, 2005(22):93-100.
- [51] Kenneth E. Hild II. Feature Extraction Using Information-Theoretic Learning. IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 28, No. 9, September 2006.

[52] Margherita Berardi, Donato Malerba, Marcella Attimonelli. Mining Information Extraction Models for HmtDB Annotation. Sixth IEEE International Conference on Data Mining - Workshops (ICDMW'06), 2006.

[53] Lipika Dey, Muhammad Abulaish, Jahiruddin, Gaurav Sharma. Text Mining Through Entity-Relationship Based Information Extraction. IEEE/WIC/ACM International Conferences on Web Intelligence and Intelligent Agent Technology-Workshops, 2007, 177-180.

[54] XiaoLing Wang, DeJun Mu, Jun Fang. Improved Automatic Keyphrase Extraction by Using Semantic Information. International Conference on Intelligent Computation Technology and Automation, 2008:1061-1065.

[55] Hiep Phuc Luong, Susan Gauch, Qiang Wang. Ontology Learning Through Focused Crawling and Information Extraction. International Conference on Knowledge and Systems Engineering, 2009:106-112.

[56] Tao Xie, Shengsheng Shi, Fuliang Quan. Research on Complex Structure-Oriented Accurate Web Information Extraction Rules [M]. Progress in Informatics and Computing (PIC), 2010 IEEE International Conference, 2010(1): 312-316.

[57] Shady Shehata, Fakhri Karray, Mohamed S. Kamel. An Efficient Concept-Based Mining Model for Enhancing Text Clustering[J]. IEEE Transactions on Knowledge and Data Engineering, 2010(22):1360-1371.

[58] Christopher Baker. Extraction and Grounding of Protein Mutations via Semantic Integration of Text and Sequence Information. International Conference on Advanced Information Networking and Applications, 2011:556-563.

[59] Christina Feilmayr. Text Mining-Supported Information Extraction[M]. 2011 22nd International Workshop on Database and Expert Systems Applications, 2011:217-221.

[60] Ning Zhong, Yuefeng Li, Sheng-Tang Wu. Effective Pattern Discovery for Text Mining[J]. IEEE Transacons on Knowledge and Data Engineering, 2012, (24):30-44.

[61] Benjamin Adrian, Jorn Hees, Ludger van Elst, Andreas Dengel. iDocument: Using Ontologies for Extracting and Annotating Information from Unstructured Text[C]. Berlin: Springer, 2009:249-256.

[62] M. Saravanan, B. Ravindran, S. Raman. Improving legal information Retrieval Using an Ontological Framework[C]. Artif Intell Law, 2009(17):101-124.

[63] 王国印. 科技文献语义标注系统研究[D]. 武汉: 华中科技大学, 2009.

[64] Freitag D, Kushmerick N. Boosted Induction[C]. In Proceedings of 17th National Conference on Artificial Intelligence, 2000, 577-583.

[65] Popov B, Kirilov A, et al. Kim-Semantic Annotation Platform[C]. In Proceedings of 2nd International Semantic Web Conference(ISWC' 2003). Florida, USA:833-849.

[66] Alani H, Kim S, Millard D, et al. Automatic Ontology-Based Knowledge Extraction from Web Documents[J]. Intelligent Systems, 2003, 18(1):13-21.

[67] Leonard T, Glaser H. Large Scale Acquisition and Maintenance from the Web without Source Access[C]. Workshop 4, Knowledge Markup and Semantic Annotation, K-CAP 2001:97-101.

[68] Kogu P, Holmes W. AeroDaml: Applying Information Extraction to Generate DAML Annotations from Web Pages[C]. In the Proceedings of K-CAP'2001 Workshop Knowledge Markup & Semantic Annotation, Canada:2001.

[69] Ciravegna F. An Adaptive Algorithm for Information Extraction from Web-related Texts[C]. In Proceedings of the IJCAI'2001 Workshop on Adaptive Text extraction and Mining Held in Conjunction with 17th International Joint Conference on Artificial Intelligence (IJCAI'2001). Seattle, USA. 2001:1251-1256.

[70] Vargas-Vera M, Motta E, Domingue J, Lanzoni M, Stutt A, Ciravegna F. MnM: Ontology Driven Semi-automatic and Automatic Support for Semantic Markup[C]. In Proceedings of the 13th International Conference on Knowledge Engineering and Management (EKAW'2002). Siguenza, Spain: 2002, 379-391.

[71] Ciravegna F, Dingli A, Iria J, Wilks Y. Multi-Strategy Definition of Annotation Services in Melita[C]. In Proceedings of ISWC'2003 Workshop on Human Language Technology for the Semantic Web and Web Services, 2003, 97-107.

[72] Benjamins R, et al. 2003. Esperanto Application: Service Provision of Semantic Annotation, Aggregation, Indexing and Routing of Textual, Multimedia and Multilingual Web content[C]. In the Proceedings of Workshop WIAMIS, 2003.

[73] Martin P, Eklund P. Embedding Knowledge in Web Documents[C]. In Proceedings of the 8th International World Wide Web Conference (WWW' 1999). Toronto, 1999, 1403-1419.

[74] Freitag D. Information extraction from HTML: Application of a General Machine Learning Approach[C]. In Proceedings of the 15th Conference on Artificial Intelligence (AAAI' 98), 517-523.

[75] Cortes C, Vapnik V. Support Vector Networks[J]. Machine Learning, 1995, 20:273-297.

[76] Collins M. Discriminative Training Methods for Hidden Markov Models: Theory and Experiments with Perceptron Algorithms[J]. In Proceedings of the Conference on Empirical Methods in NLP, 2002.

[77] Dill S, Tomlin J, et al. SemTag and Seeker: Bootstrapping the Semantic Web Via Automated Semantic Annotation[C]. Proc of the 12th Int'l Conf on World Wide Web. New York: ACM, 2003:178-186.

[78] Reeve L. Integrating Hidden Markov Models into Semantic Web Annotation

Platforms[J]. Technique Report, 2004.

[79] Lawrence S, Giles C L. Accessibility of Information on the Web[J]. Nature, 1999, 400:107-109.

[80] 袁柳等. 基于本体的 Deep Web 数据标注[J]. 软件学报, 2008, 19(2): 237-245.

[81] Handschuh S, Volz R, Staab S. Annotation for the Deep Web. IEEE Intelligent Systems, 2003, 18(5): 42-48.

[82] Jianming Li, Lei Zhang, Yong Yu. Learning to Generate Semantic Annotation for Domain Specific Sentences[C]. In Proceedings of the Knowledge Markup and Semantic Annotation Workshop in K-CAP'2001. Victoria, BC. 2001.

[83] Philipp Cimian, Siegfried Handschuh, Steffen Staab. Gimme' The Context: Context Driven Automatic Semantic Annotation with C-PANKOW[C]. WWW.2005, Chiba, Japan.

[84] Reeve L, Han H. Survey of Semantic Annotation Platforms[C]. In Proceedings of the 2005 ACM Symposium on Applied Computing. New Mexico, USA, 2005, 1633-1638.

[85] 戴维民等. 义网信息组织技术与方法[M]. 上海: 学林出版社, 2008.

[86] Atanas Kiryakov, Borislav Popov, Ivan Terziev, Dimitar Manov, Damyan Ognyanoff. Semantic Annotation, Indexing, and Retrieval[C]. 2011.

[87] Hammond B, Sheth A, Kochut K. Semantic Enhancement engine: a Modular Document Enhancement Platform for Semantic Applications over Heterogeneous Content[M]. in Real World Semantic Web Applications. V. Kashyap and L. Shklar. Eds., IOS Press, 2002, 29-49.

[88] 高琦. 基于 Bootstrapping 的本体标注方法研究[D]. 四川: 重庆大学, 2010.

[89] Sushain Pandit. Ontology-Guided Extraction of Structured Information from Unstructured Text: Identifying and Capturing Complex Relationships[D]. 2010.

[90] Theresia Gschwandtner, et al. Easing Semantically Enriched Information Retrieval An Interactive Semi-Automatic Annotation System for Medical Documents[J]. International Journal of Human-Computer Studies, 2010, 68: 370-385.

[91] 戚欣, 肖敏, 孙建鹏. 基于本体案例库的自动语义标注[J]. 计算机应用研究, 2011, 28 (5): 1742-1747.

[92] 常平梅. 一种多本体支持的语义标注模型的研究[D]. 大连: 大连海事大学, 2010.

[93] 李海芳, 等. 多维优化案例推理检索算法研究[J]. 计算机工程与技术, 2008, 44 (25): 157-160.

[94] 李锋刚, 倪志伟, 杨善林, 等. 案例推理中属性约简及其性能评价[J]. 清华大学学报 (自然科学版), 2001, 46 (S1): 1025-1029.

[95] 屈喜龙. 一个基于 AHP 的 CBR 系统实例检索模型[J]. 计算机应用研究, 2005: 33-38.

[96] 李锋, 魏莹. 分布式环境下基于语义相似的案例检索[J]. 计算机工程, 2007, 33 (9): 28-30.

[97] 杨奕飞, 瞿元新. 基于 CBR 和 RBR 的航天测控设备故障诊断专家系统[J]. 飞行器测控学报, 2005, 24 (3): 23-31.

[98] Rodriguez M Andrea, Egenhofer Max J. Determining Semantic similarity among Entity Classes from Different Ontologies [J]. Knowledge and Data Engineering, 2003, 15(2): 442-456.

[99] 王湘云. 一阶谓词逻辑在人工智能知识表示中的应用[J]. 重庆工学院学报 (社会科学版), 2007, 21 (9): 69-71.

[100] 何连双. 司法判决中的逻辑模式[J]. 司法世界, 2009 (10): 159-160.

- [101] 吴强. 语义 web 中以描述逻辑为本体语言的推理[J]. 计算机工程与应用, 2003, 33: 30-33.
- [102] 毛海飞. 自动推理技术在法律框架网络语义检索系统中的应用[D]. 山西大学硕士学位论文, 2011.
- [103] F. Baader, D. Calvanese, D. McGuinness, D. Nardi, P. Patel-Schneider. The DL Handbook[M]. Cambridge University Press, 2003.
- [104] Gruber T R. Ontolingua: A Translation Approach to Portable Ontology Specifications[J]. Knowledge Acquisition, 1993, 5 (2) : 199-220.
- [105] Lenat, Guha. Building Large Knowledge-Based System: Representation and Inference in the CYC Project [M]. AddisonWesley Press, 1990.
- [106] MacGregor. Inside the LOOM Classifier [J]. SIGART Bulletin, 1991, 2 (3): 70-76.
- [107] 胡鹤. 本体方法及其时空推理应用研究[D]. 吉林大学博士学位论文, 2004.
- [108] 王春丽. 试论法律应用中推理的可废止性[D]. 西南大学硕士学位论文, 2011.
- [109] 官礼合. 面向法律领域案例归类判决系统的研究与实践[D]. 重庆交通学院硕士学位论文, 2004.
- [110] 郑永和, 寇应展, 张伟君. 基于模糊推理的案例检索技术研究[J]. 四川兵工学报, 2009, 30 (7): 122-24.
- [111] 左春荣. 基于定性推理理论和 Multi-Agent 的群决策支持系统研究[D]. 合肥工业大学博士学位论文, 2008.
- [112] 於兴中. 法律的形式与法律推理[J]. 法律方法与法律思维, 2003: 93-106.
- [113] 徐英杰. 基于范例推理的刑事判案咨询系统的设计与实现[D]. 中国科学院研究生院硕士学位论文, 2007.

[114] Wyner A, Bench-Capon. Argument Schemes for Legal Case-Based Reasoning[J]. Mommers L (eds) Legal Knowledge and Information Systems, 2007: 139-149.

[115] Brueninghaus S, Ashley KD. Combining Model-Based and Case-based Reasoning for Predicting the Outcomes of Legal Cases[J]. Lecture Notes in Artificial Intelligence, 2003, 2689:65-79.

[116] Trevor Bench-Capon, Henry Prakken. Giovanni Sartor Argumentation in Legal Reasoning[M]. Argumentation in Artificial Intelligence, 2009: 363-382.

[117] Jim Prentzas, Ioannis Hatzilygeroudis. Categorizing Approaches Combining Rule-Based and Case-Based Reasoning[J]. Expert Systems, 2007, 24 (2): 97-122.

[118] C.R.Marling, G.J.Petot, L.S.Sterling. Integrating Case-Based and Rule-Based Reasoning to Meet Multiple Design Constraints[J]. Computational Intelligence, 1999, 15(3): 308-322

[119] 丁文严. 案例指导制度背景下人民法院案例系统的建立[J]. 法律适用, 2013, 1: 88-92.

[120] Perez A G, Benjamins V R. Overview of Knowledge Sharing and Reuse Components: Ontologies and Problem Solving Methods[C]. Proceedings of the IJCAI-99 Workshop on Ontologies and Problem Solving Methods(KRR5), Stockholm V R, Benjamins B, Chandrasekaran A, 1999:1-15.

[121] Grigoris Antoniou, Enrico Franconi. Introduction to Semantic Web Ontology Languages[J]. Reasoning Web, 2005, 3564:1-21.

[122] 余贵清, 张永安. 审判案例自动抽取与标注模型研究[J]. 现代图书情报技术, 2013, 234 (6): 23-29.

[123] 余贵清, 张永安. 基于本体的刑事审判案例推理方案研究[J]. 图书情报工作, 2014, 58 (13): 118-124.